

Máster en Matemática Industrial

Curso 2020 – Curso 2021

Proyecto Fin de Máster

Detección de Fallos en Compresores con Análisis de Vibraciones

Elba García Hermida

Fecha presentación: 20/07/2021

Tutor académico: Andrés Prieto Aneiros

Empresa: COFRICO

Tutor empresa: Jaime Herrero Paz

Índice general

Resumen	III
Abstract	IV
1. Introducción	1
1.1. Presentación de la empresa	1
1.2. Planteamiento y motivación del problema	2
2. Modelización del compresor de tornillo	3
2.1. Modelo mecánico simplificado	5
2.2. Obtención de los términos fuente	8
2.3. Resolución del sistema de ecuaciones diferenciales	10
3. Técnicas de clasificación	15
3.1. Método KMEANS	16
3.1.1. Algoritmo KMEANS++	17
3.1.2. Método KMEDOIDS	17
3.2. Método CLARANS	19
3.3. Método GMEANS	23
4. Clasificación de datos manufacturados	27
4.1. Clasificación con clusters con el mismo número de elementos	27
4.1.1. Clasificación según los coeficientes de rigidez axial de los rotores	28
4.1.2. Clasificación según los coeficientes de rigidez de la lubricación del aceite	29
4.1.3. Clasificación según los coeficientes de rigidez axial de los rotores y de la lubricación del aceite	31

4.1.4. Clasificación según los coeficientes de rigidez axial y el coeficiente de rigidez torsional conjunta	35
4.2. Clasificación con un cluster con muchos menos elementos . . .	37
4.3. Clasificación con un único cluster	41
4.4. Clasificación usando datos manufacturados con ruido	43
5. Clasificación de datos reales	47
5.1. Clasificación de nuevos datos experimentales	54
6. Conclusiones y futuro trabajo	59
Bibliografía	61

Resumen

El presente Trabajo de Fin de Máster tiene como objetivo principal estudiar si es posible determinar el estado de funcionamiento de un compresor de tornillo de un ciclo frigorífico. La empresa que ha propuesto el problema, COFRICO, tiene gran interés en la resolución de este problema, pues le permitiría mejorar el mantenimiento predictivo de sus instalaciones.

Con este fin se desarrollará un modelo matemático simplificado del compresor que nos permitirá obtener datos manufacturados que podremos clasificar mediante distintos métodos de clasificación, todos ellos implementados en *Python*. Una vez comprobada la eficacia de los algoritmos de clasificación con los datos manufacturados se procederá a aplicarlos a los datos reales proporcionados por la empresa. Como los datos reales disponibles proceden de un compresor en un buen estado de funcionamiento se introducirán artificialmente datos de malfuncionamiento.

Se comprobará así que es posible detectar un posible funcionamiento incorrecto del compresor y valdrá como punto de partida para que la empresa pueda seguir desarrollando el método propuesto en el trabajo.

Abstract

The main objective of this Master's Thesis is to classify the operating status of a screw compressor in a refrigeration cycle using vibration data. The company that has proposed the problem, COFRICO, has great interest on this problem since it could increase the performance of the predictive maintenance of its facilities.

With this aim, a simplified mathematical model of the compressor have been developed. Thanks to this mechanical model, some manufactured data has been generated and used to check different classification methods, all of them implemented in Python. Once the accuracy of the classification algorithms with the manufactured data has been verified, they have been applied to real experimental data provided by the company. As the actual available data come from a compressor in good operating status, additional malfunction data have been artificially created.

In conclusion, it has been verified that it is possible to detect a possible malfunction of the compressor under some particular conditions. These findings will be a starting point for the company to further develop the methodology proposed in this project.

Capítulo 1

Introducción

1.1. Presentación de la empresa

COFRICO nace en 1985 en Burela, Lugo, con el objetivo de dar respuesta a las necesidades del mundo empresarial en el campo de la refrigeración y la climatización. Desde su creación siguió expandiéndose hasta convertirse en una de las principales empresas del país en este ámbito. COFRICO apuesta por la eficiencia energética y la reducción del impacto de la industria en el medio ambiente, siendo pionera en las instalaciones con gas CO_2 refrigerante.

La empresa brinda a sus clientes un mantenimiento preventivo y predictivo de las instalaciones, lo que se traduce en un mayor ahorro y eficiencia, llevado a cabo gracias al sistema *CofriView365*, que integra de forma centralizada los datos de todas las variables energéticas y de proceso de la industria. Su área principal de negocio es la hostelería y la restauración industrial. Gracias al sistema *CofriView365* han conseguido reducir los costes y consumos hasta un 50 % (para más información ver el sitio web de la empresa [12]). Actualmente este sistema se encuentra en desarrollo hacia el mantenimiento predictivo.

La innovación, la eficiencia energética, el compromiso con el medioambiente y la calidad de servicio son los principales valores de la empresa, y se reflejan en el continuado afán por buscar nuevas soluciones y ofrecer el mejor servicio a sus clientes.

1.2. Planteamiento y motivación del problema

En un ciclo frigorífico básico, el refrigerante, por ejemplo amoníaco, entra en el sistema a través del compresor en estado gas. En el compresor se aumentará su presión, aumentando su temperatura. Una vez en el condensador comienza la fase de enfriamiento del gas hasta alcanzar su estado líquido. Finalmente se reducirá la temperatura y la presión de nuevo para enfriar la cámara, que es donde se realiza el intercambio de calor.

En este proceso, la componente más importante, más costosa y que genera el mayor consumo es el compresor. Es por ello que la empresa está interesada en poder predecir un posible mal funcionamiento del compresor mediante el estudio de sus vibraciones. Los compresores, como toda maquinaria industrial, produce vibraciones, las cuales son un indicador del buen o mal funcionamiento del aparato. Por ello, se pretende simular el comportamiento del compresor para obtener medidas de vibraciones, con el objetivo de entrenar un algoritmo de aprendizaje que permita detectar las posibles disfuncionalidades del compresor

El mayor problema reside en ser capaces de modelar el comportamiento de los compresores y poder predecir la vida útil de éste, permitiendo adelantarse a su posible fallo y reducir tiempos de reparación (con pedidos anticipados de piezas). De esta manera, se pretende poder realizar un mantenimiento predictivo eficaz que, combinado con el sistema *CofriView365*, permita garantizar al cliente el correcto funcionamiento del sistema el mayor tiempo posible.

La empresa ha proporcionado datos de vibraciones de uno de sus compresores a lo largo de varias semanas, así como el estado de éste durante ese tiempo. El principal reto al que nos enfrentamos es que dicho compresor no presenta fallos, pues estos se manifiestan principalmente a largo plazo. Como veremos en los próximos capítulos se han empleado dos enfoques diferentes para evitar este inconveniente y obtener resultados satisfactorios sobre los que la empresa pueda seguir trabajando. El primero se llevó a cabo en una primera evaluación del problema, cuando no se disponían de datos de vibraciones, se ha modelizado el comportamiento del compresor para generar una gran cantidad de datos de vibraciones generados a partir de un gemelo digital (que incluye un modelo mecánico simplificado) que nos permitan preparar los algoritmos de clasificación para su uso en el caso de unos datos reales. En el segundo enfoque se clasificarán los datos reales introduciendo artificialmente unos datos de vibraciones que pueden asociarse a un comportamiento defectuoso del aparato.

Capítulo 2

Modelización del compresor de tornillo

El fenómeno principal que se pretende analizar son las vibraciones generadas en el compresor y cómo estas se asocian con el fallo del aparato. Estas vibraciones dependen, en gran medida de cómo es el funcionamiento del compresor de tornillo y cómo interaccionan sus componentes entre si. A pesar de las diferencias estructurales, los distintos compresores de tornillo comparten los mismos elementos principales:

- Rotores: Conformados por un rotor macho y un rotor hembra que al girar a altas velocidades crean unas cámaras a través de las cuales el gas se aspira, comprime y descarga. Son las piezas principales del modelo.
- Cilindros de compresión: Aloja a los rotores. El gas aspirado pasa entre el cilindro y los rotores. No aparecen en el modelo.
- Filtros de aire: Se sitúa en el interior de la válvula de apertura y recoge polvo, partículas y humedad, evitando que se dañe el interior de la máquina. No se tendrán en cuenta en el modelo simplificado.
- Filtros de aceite: En los compresores lubricados por aceite (como es el caso estudiado) se sitúan en las paredes de la cámara giratoria y cerca de las válvulas de descarga. Su objetivo es eliminar posibles restos de aceite del aire comprimido y garantizar un enfriamiento correcto del aparato. No se considerarán en el modelo.
- Cojinetes: Se sitúan a ambos extremos de los rotores para asegurar los rotores. En el modelo se simplifican como muelles.

4 CAPÍTULO 2. MODELIZACIÓN DEL COMPRESOR DE TORNILLO

- Válvula de succión: Generalmente se encuentra en la parte superior del compresor y se abrirá para permitir la entrada del aire. No aparecerán en el modelo simplificado pero se disponen de datos de presión y temperatura, como se verá en el capítulo 5.
- Válvula de descarga: Se encuentran en el extremo opuesto a las válvulas de succión y finalizan el ciclo de compresión. No aparecerán en el modelo simplificado pero también se disponen de datos de presión y temperatura.
- Motor: Proporciona la potencia necesaria para el funcionamiento del compresor. Aparecerá en el modelo.
- Tanques de almacenamiento: Almacenan el aire comprimido garantizando que éste mantiene su presión. No se tienen en cuenta en el modelo.
- Válvula corredera: Permite controlar la capacidad del compresor, cuando está en posición cerrada el compresor funciona a plena capacidad y a medida que se desliza axialmente a la posición abierta se reduce la capacidad. Aunque no se contempla en el modelo, en el capítulo 5 se utilizará como indicador de que el aparato está en funcionamiento.

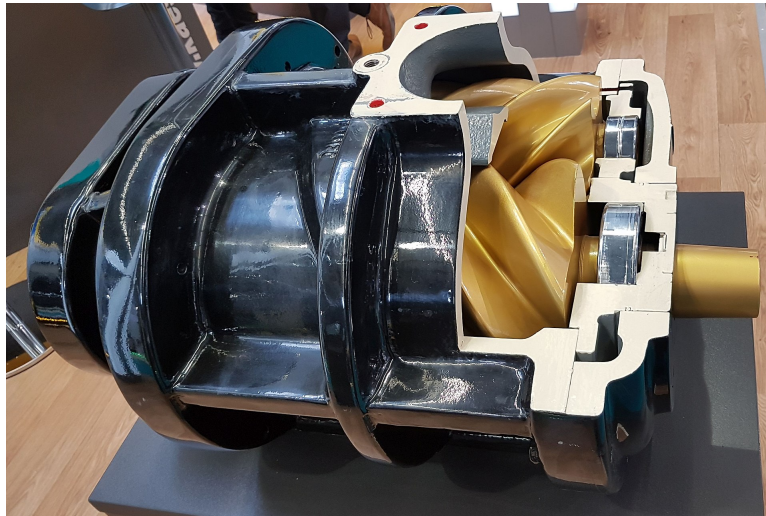


Figura 2.1: Compresor de tornillo. Imagen de [1]

Es necesario tener claro cómo funciona el compresor de tornillo, que es una pieza indispensable en los sistemas de refrigeración ya que es el encargado de comprimir el refrigerante en estado gas, elevando su temperatura y presión. Su funcionamiento se basa en el desplazamiento del gas a través de cámaras que se crean con el giro simultáneo y en sentido contrario de dos tornillos, uno macho y uno hembra. El gas llena los espacios aumentando su presión a medida que disminuye el volumen de dichas cámaras. En el extremo opuesto se encuentra una tobera de salida, por donde sale el gas ya comprimido. En la Figura 2.1 se puede ver una imagen de los rotores en un compresor de tornillo similar a los utilizados por la empresa.

2.1. Modelo mecánico simplificado

Para el estudio de las vibraciones se considerará una simplificación del sistema, formado unicamente por el rotor macho, impulsado por el motor eléctrico, y el rotor hembra, ambos con sendos cojinetes de empuje, como se muestra en la Figura 2.2.

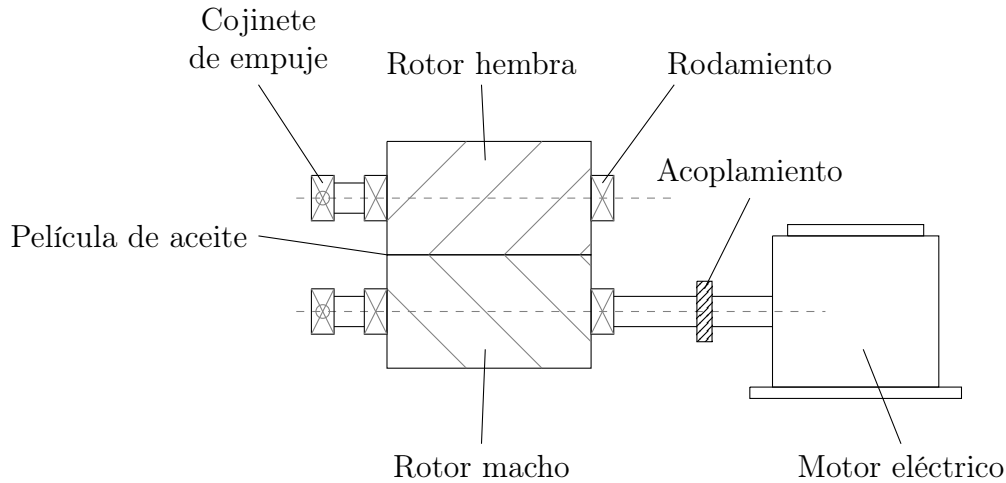


Figura 2.2: Sistema simplificado del compresor de tornillo. Sistema estudiado en [13].

Se realiza la misma simplificación del sistema que se lleva a cabo en [13]. Se considera así el compresor como un sistema mecánico sometido a una fuerza exterior que puede variar con el tiempo y donde se asumen las siguientes hipótesis:

6 CAPÍTULO 2. MODELIZACIÓN DEL COMPRESOR DE TORNILLO

- Los rotores del compresor y el rotor del motor son cuerpos rígidos, mientras que el acoplamiento y el eje de transmisión se consideran con un coeficiente de rigidez torsional conjunta k_3 .
- Los cojinetes de empuje se consideran como muelles con coeficientes de rigidez axial k_1 y k_2 .
- Las interacciones entre los lóbulos de los rotores del compresor de tornillo se pueden simplificar a las de engranajes helicoidales. Se utilizará en el análisis el ángulo de la hélice δ_p .
- La película de aceite entre los rotores macho y hembra proporciona coeficientes de rigidez radial $k_{0,r}$, axial $k_{0,a}$ y circunferencial $k_{0,t}$.
- La influencia de la deformación por flexión del eje es despreciable.
- La amplitud del par vibratorio en los rotores es menor que el par medio transmitido. Esto es cierto en la mayoría de los casos y conlleva a que la respuesta del ensamblaje del rotor no afecta a los cálculos de la frecuencia (véase el capítulo 7 en [13] para más detalles).

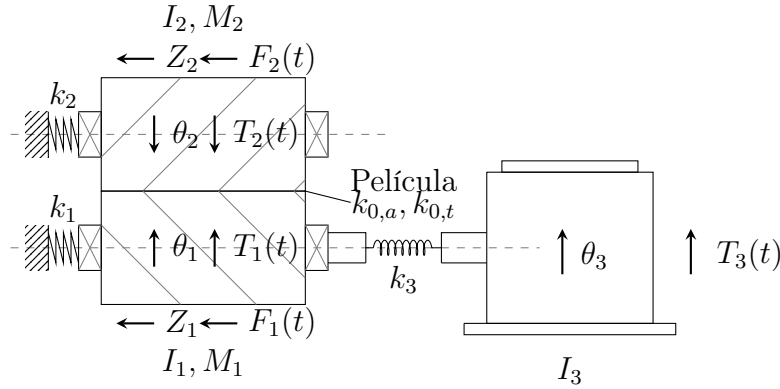


Figura 2.3: Sistema idealizado del compresor de tornillo. Imagen de [13].

Para modelar el comportamiento mecánico del compresor denotamos por M_1 y M_2 la masa, por I_1 y I_2 a los segundos momentos de área, por $F_1(t)$ y $F_2(t)$ a las fuerzas axiales aplicadas, por $T_1(t)$, $T_2(t)$ y $T_3(t)$ a los torques, por Z_1 y Z_2 a los desplazamientos axiales y por θ_1 , θ_2 y θ_3 a los movimientos

angulares. Los subíndices 1, 2, 3 se refieren con el rotor macho, el rotor hembra y la fuerte motriz, respectivamente.

De esta idealización se extrae un modelo de vibraciones

$$\mathcal{M}\ddot{\mathbf{x}} + \mathcal{K}\mathbf{x} = \mathbf{f}, \quad (2.1)$$

donde la matriz de masa es

$$\mathcal{M} = \begin{pmatrix} M1 & 0 & 0 & 0 & 0 \\ 0 & M2 & 0 & 0 & 0 \\ 0 & 0 & I1 & 0 & 0 \\ 0 & 0 & 0 & I2 & 0 \\ 0 & 0 & 0 & 0 & I3 \end{pmatrix}, \quad (2.2)$$

el vector de incógnitas es

$$\mathbf{x}(t) = (Z_1(t), Z_2(t), \theta_1(t), \theta_2(t), \theta_3(t))^T, \quad (2.3)$$

el vector de fuerzas y momentos es

$$\mathbf{f}(t) = (F_1(t), F_2(t), T_1(t), T_2(t), T_3(t))^T, \quad (2.4)$$

y la matriz de rigidez es

$$\mathcal{K} = \begin{pmatrix} k_1 & -k_{0,a} & r_{p,1}k_{0,t} \tan \delta_p & r_{p,2}k_{0,t} \tan \delta_p & 0 \\ -k_{0,a} & k_2 + k_{0,a} & r_{p,1}k_{0,t} \tan \delta_p & -r_{p,2}k_{0,t} \tan \delta_p & 0 \\ -r_{p,1}k_{0,t} \tan \delta_p & r_{p,1}k_{0,t} \tan \delta_p & k_3 + r_{p,1}^2 k_{0,t} & -r_{p,1}r_{p,2}k_{0,t} & -k_3 \\ r_{p,2}k_{0,t} \tan \delta_p & -r_{p,2}k_{0,t} \tan \delta_p & -r_{p,1}r_{p,2}k_{0,t} & r_{p,2}^2 k_{0,t} & 0 \\ 0 & 0 & -k_3 & 0 & k_3 \end{pmatrix}. \quad (2.5)$$

Por tanto, para resolver del sistema de ecuaciones diferenciales (2.1) es necesario conocer los coeficientes de rigidez $k_1, k_2, k_3, k_{0,r}, k_{0,a}$ y $k_{0,t}$, así como el ángulo de la hélice δ_p , y los radios de los rotores $r_{p,1}$ y $r_{p,2}$. Aunque obtener el ángulo de la hélice y el radio de los rotores es posible, conocer los coeficientes de rigidez no es una tarea tan fácil. Se podría tratar de utilizar la teoría de contacto de Hertz tal y como se hace en [6], según la cual

$$k_{0,r} = 2E_s \left(\frac{3F_n \rho_s}{4E_s} \right)^{1/3}, \quad (2.6)$$

donde $E_s = E_1 E_2 / (E_1 + E_2)(1 - \nu^2)$ es el módulo elástico sintético, siendo E_1 y E_2 los módulos elásticos de los cuerpos en contacto, ν el ratio de Poisson.

$\rho_s = \rho_1 \rho_2 / (\rho_1 + \rho_2)$ el radio de curvatura sintético, con ρ_1 y ρ_2 los radios de curvatura de los cuerpos en el punto de contacto. Sin embargo, en este caso dado que los componentes sólidos son asumidos rígidos y lubricados con aceite, esta expresión no podría utilizarse directamente.

De hecho, si se tiene en cuenta que los rotores están en constante movimiento, los radios de curvatura de las superficies de contacto variarán con el tiempo, haciendo a su vez que los coeficientes de rigidez dependan del tiempo. No es al único problema al que nos enfrentamos al querer calcular estos coeficientes, pues dependerán también del estado del aceite y el desgaste entre otras variables.

Sin embargo para clasificar el estado de un sistema mecánico no nos interesa crear un modelo matemático detallado que se ajuste a la realidad de forma muy precisa, nos basta con que su comportamiento sea similar. A la hora de resolver el sistema de ecuaciones (2.1) se darán valores estimados a estos parámetros desconocidos. Como los coeficientes de rigidez dependerán del desgaste del aparato (que será un indicativo de su buen o mal funcionamiento) resolveremos el sistema de ecuaciones (2.1) para distintos valores de los coeficientes de rigidez y se comprobará si los clasificadores son capaces de distinguir curvas de vibraciones que provengan de distintos coeficientes de rigidez (y como consecuencia, de distintos estados de desgaste).

2.2. Obtención de los términos fuente

Los términos fuente (2.4) considerados, necesarios para la resolución del sistema de ecuaciones diferenciales (2.1), serán los obtenidos a partir de las estimaciones realizadas en [13], que son resultado del análisis de fuerzas a lo largo de varios ciclos de rotación del compresor.

En [13] estas funciones fuente aparecen representadas en gráficas, por lo que para poder hacer uso de ellas se digitalizaron usando alrededor de 50 puntos tratando que estos fueran equiespaciados. A partir de estos datos se construyeron sus funciones interpoladoras.

En este trabajo se ha empleado una interpolación de tipo *B-spline* cúbica con una condición de suavizado para cada función. De manera general, dada una sucesión de números estrictamente crecientes $t_0 < t_1 < \dots < t_{m-1}$ una *B-spline* es una curva paramétrica $S : [t_0, t_{m-1}] \rightarrow \mathbb{R}$ dada por una

combinación lineal de *B-splines* básicas $B_{i,n}$ de grado n (en este caso $n = 3$)

$$S(t) = \sum_{j=0}^{m-n-2} P_j B_{j,n}(t), \quad t \in [t_0, t_{m-n}], \quad (2.7)$$

donde P_j son los puntos de control o de Boor. Las $m - (n + 1)$ *B-splines* básicas se pueden definir de forma recursiva (ver [11] para más detalle)

$$B_{j,0} := \begin{cases} 1 & t_j < t < t_{j+1}, \\ 0 & \text{en otro caso} \end{cases} \quad (2.8)$$

$$B_{j,n}(t) := \frac{t - t_j}{t_{j+n} - t_j} B_{j,n-1} + \frac{t_{j+n-1} - t}{t_{j+n-1} - t_{j+1}} B_{j+1,n-1}(t), \quad (2.9)$$

con $1 \leq j \leq m - 1$ y $1 \leq n \leq 3$.

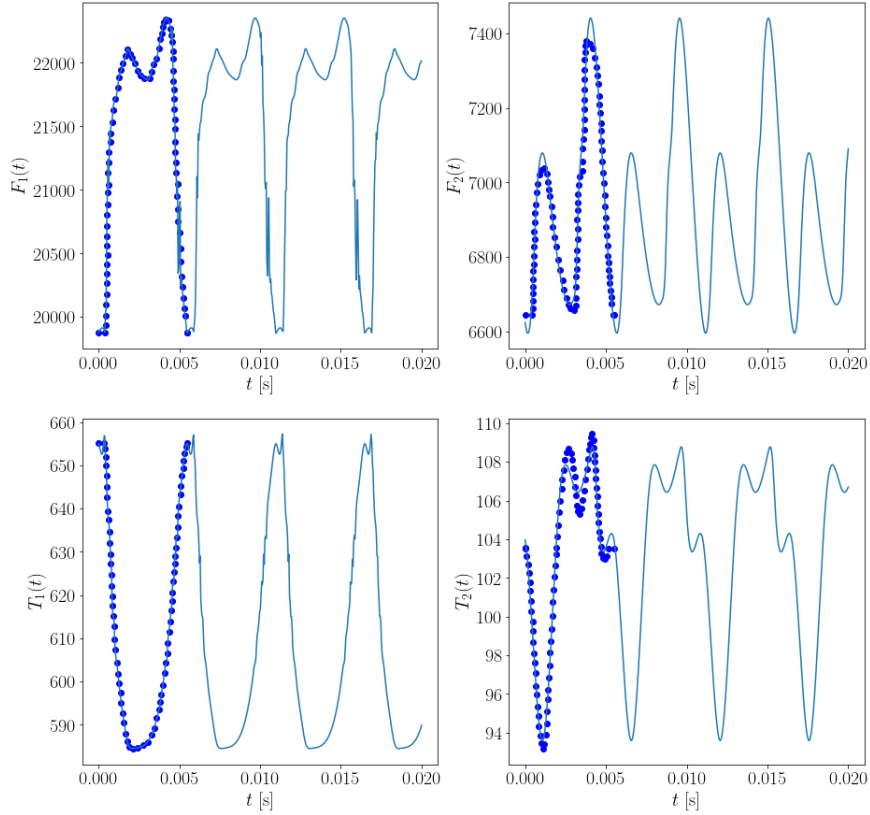


Figura 2.4: Funciones fuente: $F_1(t)$ (arriba izquierda), $F_2(t)$ (arriba derecha), $T_1(t)$ (abajo izquierda) y $T_2(t)$ (abajo derecha).

En la Figura 2.4 aparecen como puntos los datos de la digitalización de las gráficas de los cuatro términos fuente, además de las funciones periódicas resultantes de la interpolación de estos puntos.

2.3. Resolución del sistema de ecuaciones diferenciales

Es necesario resolver ahora el sistema de ecuaciones diferenciales ordinarias de orden dos que aparece en (2.1). Nótese que se trata de un sistema de cinco ecuaciones diferenciales lineales de coeficientes constantes, sin embargo, debido a que carecemos de información sobre el motor eléctrico, consideraremos que el movimiento se transmite perfectamente entre el motor y el rotor macho, reduciendo el sistema a cuatro ecuaciones diferenciales ordinarias.

Por otra parte, nótese que los coeficientes de rigidez k_1 , k_2 , $k_{0,a}$ y $k_{0,t}$ son difíciles de calcular, por lo que para poder resolver el sistema se les establecerán valores ficticios que no han podido ser validados experimentalmente. En capítulos posteriores se realizará la clasificación de las vibraciones en función de la variación de estos valores.

Debido al carácter periódico de las funciones fuente en un intervalo $(0, T)$, es posible emplear transformadas discretas de Fourier para la resolución del sistema de manera muy sencilla. Supondremos que el vector de incógnitas $\mathbf{x}(t)$ se puede aproximar mediante una serie de Fourier truncada de la forma

$$\mathbf{x}(t) = \sum_{n=-N/2}^{N/2} e^{\frac{-2\pi i t n}{NT}} \hat{\mathbf{x}}_n, \quad (2.10)$$

donde $\hat{\mathbf{x}}_n$ son los vectores con las amplitudes de Fourier de cada incógnita (donde estamos suponiendo que N es par). De forma análoga, el término fuente puede aproximarse como

$$\mathbf{f}(t) = \sum_{n=-N/2}^{N/2} e^{\frac{-2\pi i t n}{NT}} \hat{\mathbf{f}}_n, \quad (2.11)$$

Teniendo en cuenta que al derivar con respecto al tiempo resulta

$$\ddot{\mathbf{x}}(t) = -\frac{4\pi^2}{T^2 N^2} \sum_{n=-N/2}^{N/2} n^2 e^{\frac{-2\pi i t n}{NT}} \hat{\mathbf{x}}_n, \quad (2.12)$$

y sustituyendo en la ecuación diferencial (2.1), se tiene que

$$\sum_{n=-N/2}^{N/2} \left(K - \frac{4\pi^2 n^2}{T^2 N^2} M \right) \hat{\mathbf{x}}_n e^{\frac{2\pi i t n}{NT}} = \sum_{n=-N/2}^{N/2} \hat{\mathbf{f}}_n e^{\frac{-2\pi i t n}{NT}}, \quad (2.13)$$

Puesto que los modos de Fourier constituyen una base de Hilbert del espacio $(L^2(0, T))^4$ es posible igualar los coeficientes de la ecuación (2.13) uno a uno de forma que resulta

$$\left(K - \frac{4\pi^2 n^2}{T^2 N^2} M \right) \hat{\mathbf{x}}_n = \hat{\mathbf{f}}_n, \quad \text{para } -N/2 \leq n \leq N/2 \quad (2.14)$$

obteniendo un sistema de ecuaciones lineales (con matriz diagonal), cuya resolución nos proporciona el espectro de los desplazamientos axiales y el movimiento angular. La Figura 2.5 muestra estas funciones en el caso en el que se ha considerado $M_1 = M_2 = 10.0$ kg, $I_1 = I_2 = 5.0$ kg·m, $k_1 = k_2 = 1 \cdot 10^3$ N/m, $k_{0,a} = 10.0$ N/m, $k_{0,t} = 2.0$ N/m, $k_3 = 1 \cdot 10^4$ N/m, $\delta_p = \pi/4$ rad, $r_{p,1} = r_{p,2} = 0.1$ m, escogiendo $N = 128$ y tomando como $T = 0.005$ s que se corresponde con el periodo de las funciones fuente. Nótese que para cada modo de Fourier de índice n , su frecuencia es $\frac{n}{NT}$ medida en Herzios [Hz]

Finalmente, se puede recuperar la expresión de la solución en el dominio temporal usando (2.10), obteniéndose las gráficas de la Figura 2.6.

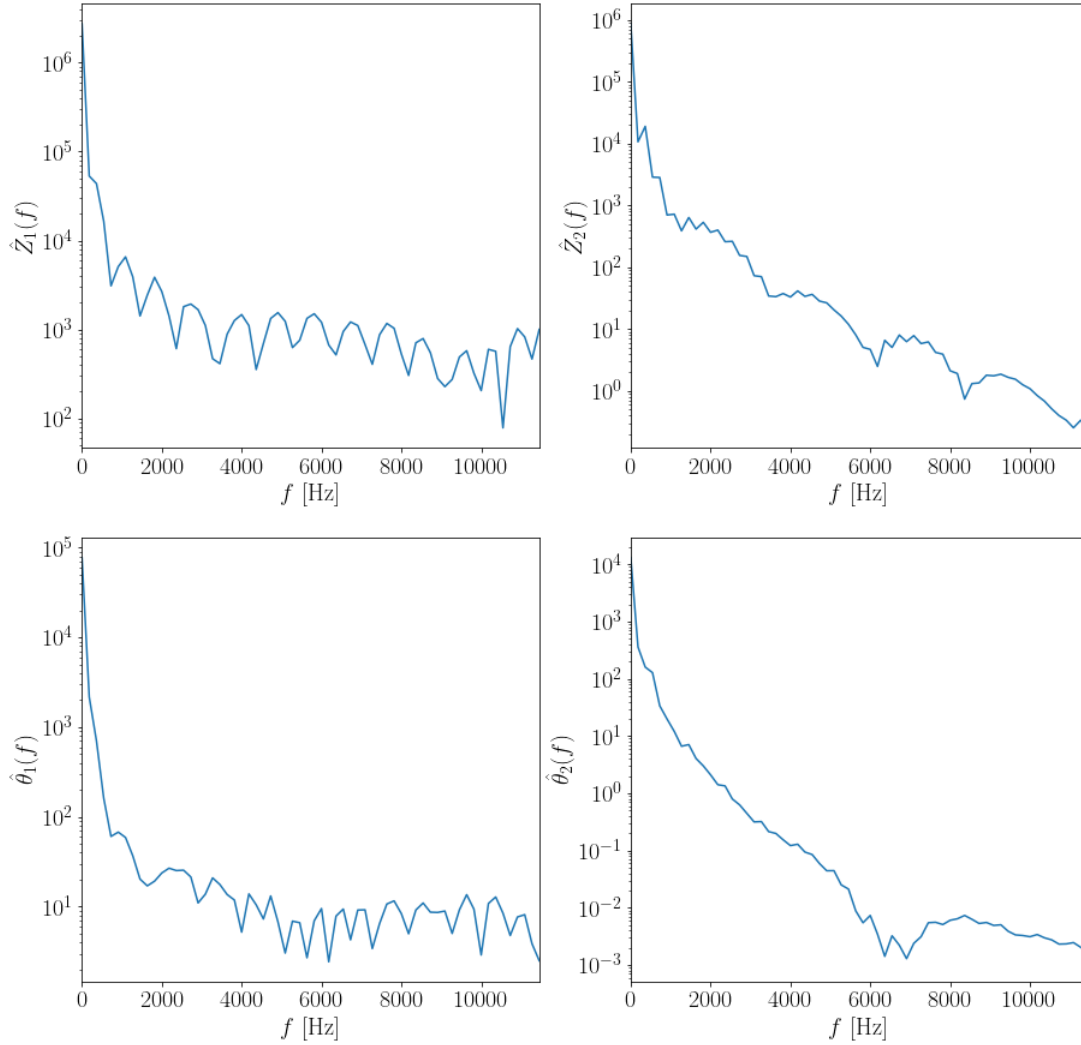


Figura 2.5: Espectro de los desplazamientos axiales y el movimiento angular: desplazamiento axial del rotor macho \hat{Z}_1 (arriba a la izquierda), desplazamiento axial del rotor hembra \hat{Z}_2 (arriba a la derecha), movimiento angular del rotor macho $\hat{\theta}_1$ (abajo a la izquierda) y movimiento angular del rotor hembra $\hat{\theta}_2$ (abajo a la derecha).

2.3. RESOLUCIÓN DEL SISTEMA DE ECUACIONES DIFERENCIALES13

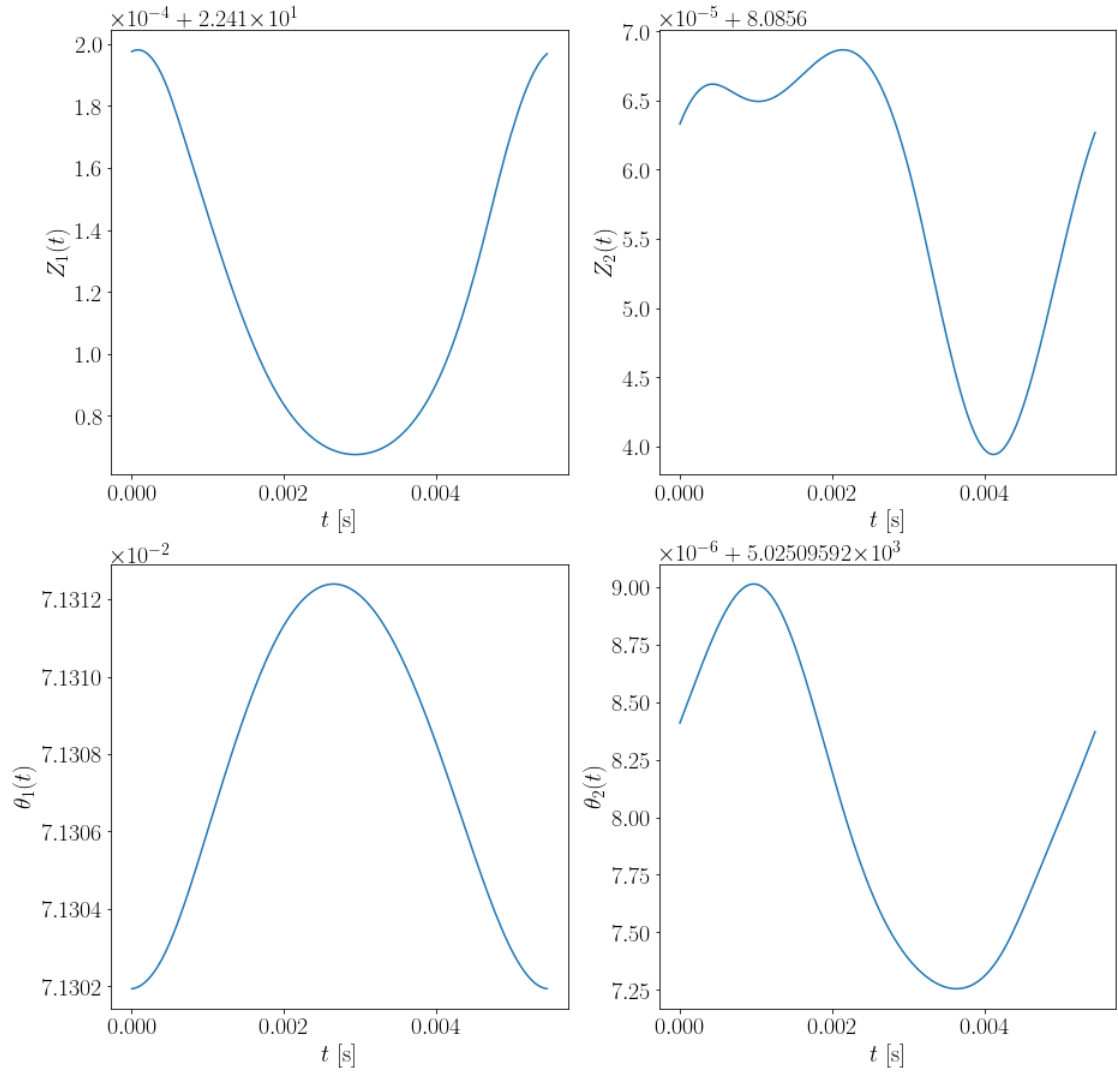


Figura 2.6: Desplazamientos axiales y movimientos angulares en función del tiempo: desplazamiento axial del rotor macho Z_1 (arriba a la izquierda), desplazamiento axial del rotor hembra Z_2 (arriba a la derecha), movimiento angular del rotor macho θ_1 (abajo a la izquierda) y movimiento angular del rotor hembra θ_2 (abajo a la derecha).

Capítulo 3

Técnicas de clasificación

En este capítulo se aplicarán algoritmos de clasificación (identificación de *clusters*) que sean capaces de detectar comportamientos distintos entre las soluciones del sistema obtenidas en el capítulo anterior. Para ello, como ya se ha mencionado, se modificarán los coeficientes de rigidez asociados al compresor k_1 , k_2 , $k_{0,a}$ y $k_{0,t}$. Estos coeficientes no sólo son difíciles de calcular, también varían con el tiempo y tienen una fuerte relación con el estado de deterioro en el que se encuentran las piezas del sistema. Es por ello que se plantea la clasificación de los resultados de las vibraciones atendiendo a los valores de estos coeficientes de rigidez. Por tanto, el algoritmo ha de ser capaz de detectar de qué tipo de coeficientes procede cada solución del sistema y clasificarla de este modo.

Los algoritmos de clasificación que se emplearán serán CLARANS, KMEANS (que se vale de KMEANS++ para encontrar los centros de los clusters) y GMEANS, todos ellos implementados en la librería de *Python* [9]. Para la descripción de estos algoritmos se usarán los conceptos de centroide y medoide

Definición 3.1 *Se define el centroide de un cluster como la media aritmética de todos los elementos que pertenecen a dicho cluster.*

Definición 3.2 *Se define el medoide de un cluster como la curva que minimiza la suma de las distancias con el resto de curvas del cluster, es decir, la curva que se encuentra en la posición media con respecto a todas las curvas del cluster.*

La principal diferencia entre un medoide y un centroide es que el medoide

debe ser una de las curvas del conjunto a clasificar mientras que el centroide no tiene por qué serlo.

3.1. Método KMEANS

El algoritmo *KMEANS* [3] es ampliamente utilizado debido a su simplicidad. Se trata de un algoritmo iterativo que trata de dividir el conjunto de datos de tamaño M en K subgrupos de intersección vacía (*clusters*). Trata de que todos los puntos de un mismo *cluster* sean lo más parecidos posibles, así como que los *clusters* sean lo más distintos entre si como sea posible. De manera general, asigna puntos a un *cluster* de forma que la suma de las distancias al cuadrado entre los datos y el centroide del *cluster* sea mínima.

El algoritmo se desarrolla en dos etapas, una en la que asigna los datos al *cluster* más cercano (*E-step*), y una en la que calcula el centroide de cada *cluster* (*M-step*).

Obsérvese que, matemáticamente la función objetivo a minimizar es la siguiente: si tenemos que identificar K *clusters* en un conjunto de M curvas, la función a minimizar J viene dada por

$$J = \sum_{i=1}^M \sum_{k=1}^K w_{ik} d(O_i, C_k)^2, \quad (3.1)$$

donde $w_{ik} = 1$ si la curva O_i pertenece al *cluster* k y $w_{ik} = 0$ en otro caso. Además C_k denota el centroide del *cluster* al que puede pertenecer O_i .

Se trata de un problema de minimización en dos partes. Primero minimizaremos J con respecto a w_{ik} , considerando C_k fijo, lo que se corresponde con el *E-step* mencionado anteriormente. A continuación minimizaremos J con respecto a C_k , fijando w_{ik} , que se identifica a su vez con el *M-step*.

De esta forma, el *E-step* consiste en encontrar w_{ik} tales que

$$\frac{\partial J}{\partial w_{ik}} = \sum_{i=1}^M \sum_{k=1}^K d(O_i, C_k)^2 = 0 \Rightarrow w_{ik} = \begin{cases} 1 & \text{si } k = \underset{1 \leq j \leq M}{\operatorname{argmin}} d(O_i, C_j)^2, \\ 0 & \text{en otro caso,} \end{cases}, \quad (3.2)$$

es decir, se asignan los elementos O_i al *cluster* más cercano según la distancia al cuadrado al centroide. Por su parte el *M-step* consiste en encontrar C_k tales

que

$$\frac{\partial J}{\partial \mu_k} = 2 \sum_{i=1}^M w_{ik}(O_i - C_k) = 0 \Rightarrow C_k = \frac{\sum_{i=1}^M w_{ik} O_i}{\sum_{i=1}^M w_{ik}}, \quad (3.3)$$

que se puede interpretar como volver a calcular el centroide de cada *cluster*, donde se ha tenido en cuenta que $d(O_i, C_k) = \|O_i - C_k\|$ es la distancia en norma ℓ^2 .

Nótese que el valor con el que se inicializan los centroides es muy importante y pueden llevar a distintas soluciones. En este caso, los centroides se inicializan mediante el algoritmo *KMEANS++*, que se explica con más detalle en la siguiente sección.

3.1.1. Algoritmo KMEANS++

El algoritmo *KMEANS++* nos permite seleccionar los centroides iniciales de manera más eficiente que simplemente tomando puntos al azar, mejorando los resultados finales de la clasificación.

Su funcionamiento es muy simple, inicialmente selecciona un centroide C_k al azar y calcula la distancia de todos los elementos del conjunto a esta curva. Si ya se han seleccionado $\tilde{K} - 1$ centroides $C_1, \dots, C_{\tilde{k}-1}$ el centroide $C_{\tilde{k}}$ se selecciona como:

$$C_{\tilde{K}} = \operatorname{argmax}_{1 \leq i \leq M} d(O_i), \quad (3.4)$$

donde $d(O_i)$ es la distancia máxima entre un elemento O_i y el centroide más alejado ya calculado

$$d(O_i) = \max_{1 \leq k \leq \tilde{K}} d(O_i, C_k)^2, \quad (3.5)$$

siendo \tilde{K} el número de centroides ya escogidos. De esta manera, el nuevo centroide será el elemento O_i que esté más alejado del resto de centroides. Este paso se repite hasta obtener los K centroides necesarios para inicializar el algoritmo *KMEANS*.

3.1.2. Método KMEDOIDS

Una vez clasificadas las curvas resultará de interés saber a qué *cluster* pertenecen nuevas curvas sin necesidad de reclasificar todos los datos. Este

proceso se realiza de modo manera muy rápida con el método *KMEDOIDS* implementado en la librería de *Python* [9], pues dispone de una función que dada una curva calcula el *cluster* más cercano.

El algoritmo *KMEDOIDS* funciona de manera idéntica al algoritmo KMEANS pero los centros de los *clusters* en lugar de ser centroides son medoides, es decir, la curva considerada como centro del *cluster* debe pertenecer al conjunto de datos.

3.2. Método CLARANS

El método CLARANS (*Clustering Large Applications based upon RAN-domized Search*) es un método de *clustering* basado en particiones.

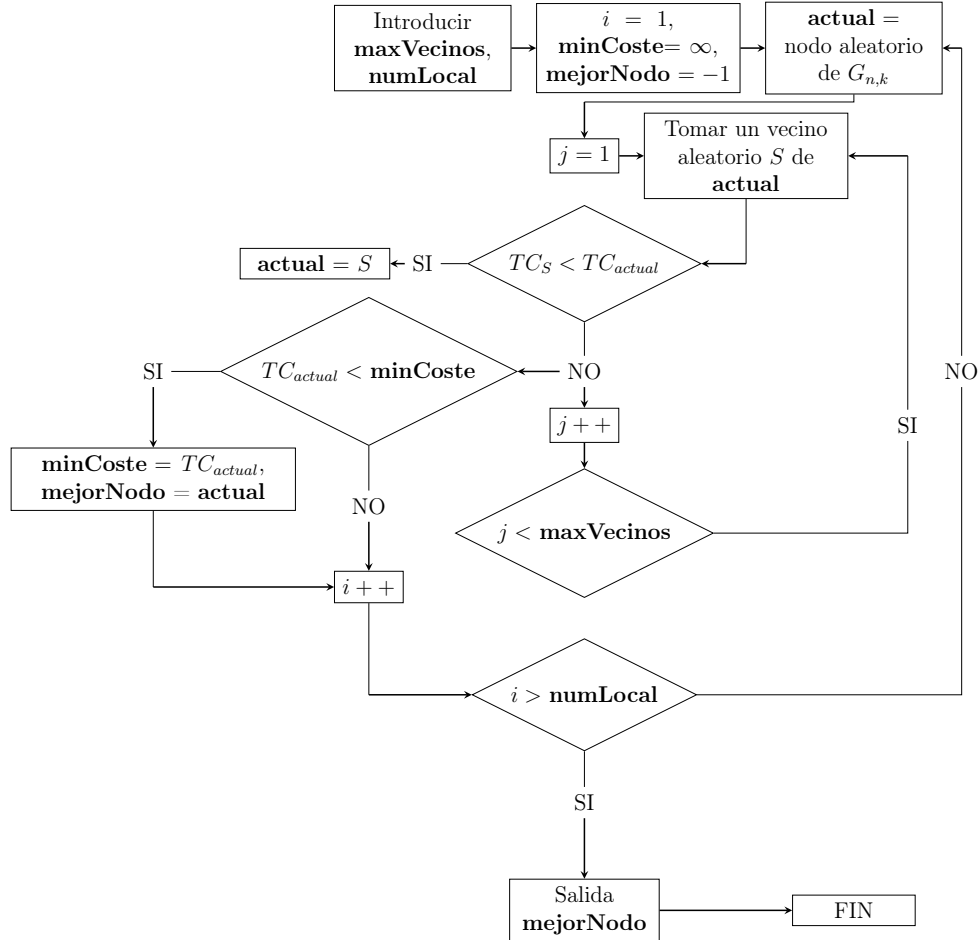


Figura 3.1: Algoritmo CLARANS. (Imagen de [4]).

El método CLARANS (*Clustering Large Applications based upon RAN-domized Search*) es un método de *clustering* basado en particiones. En el presente trabajo, su funcionamiento básico consiste en seleccionar al azar K medoides del conjunto de curvas que forma la base de datos. A continuación selecciona, también al azar, un medoide y otra curva que no sea uno de los medoides actuales y comprueba si al reemplazar uno por otro se mejora el

coste total, si es así, se intercambian.

El algoritmo requiere que se le indique un número máximo de iteraciones. En la Figura 3.1 se muestra un esquema del funcionamiento del algoritmo CLARANS.

Para precisar el funcionamiento de este algoritmo, la búsqueda de K medoides en un conjunto de M elementos se puede ver de forma abstracta como la búsqueda través de un cierto grafo, como se muestra en [8]. En este grafo, un nodo $G_{n,k}$ está representado por un conjunto de k elementos $\{O_{m_1}, \dots, O_{m_K}\}$, indicando que O_{m_1}, \dots, O_{m_K} son los medoides seleccionados. Con esta notación, los nodos del grafo son los conjuntos de K elementos de la base de datos.

Se dice que dos nodos son vecinos si únicamente difieren en un elemento, es decir, dos nodos $S_1 = \{O_{m_1}, \dots, O_{m_K}\}$ y $S_2 = \{O_{l_1}, \dots, O_{l_K}\}$ son vecinos si, y sólo si la cardinalidad de su intersección es $K - 1$. Cada nodo tendrá por tanto $K(M - K)$ vecinos. Como cada nodo representa una colección de K medoides, cada nodo se corresponderá con un *cluster*. Es por ello que a cada nodo se le puede asignar un coste (es decir, el valor de una función objetivo a minimizar), definido por la distancia total entre cada curva y el medoide de su *cluster*.

Usando la notación siguiente:

- O_m denota al medoide actual.
- O_p denota el medoide que reemplazará a O_m .
- O_j denota otra curva que no es medoide que puede ser movido o no.
- \tilde{O}_j denota al medoide más cercano a O_j , excluyendo a O_m y O_p .
- d denota la distancia entre dos elementos de la base de datos.

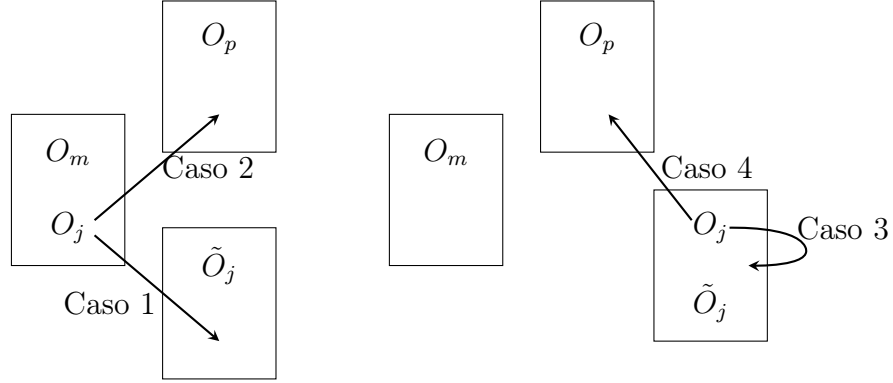


Figura 3.2: Cuatro casos posibles al reemplazar O_m por O_p . (Imagen de [8]).

Para cada curva O_j se puede calcular el coste C_{jmp} resultante de intercambiar O_m y O_p . Tenemos cuatro casos, (véase para más detalle [8]):

- Caso 1. Supongamos que O_j pertenece al *cluster* representado por O_m . Además, supongamos que O_j es más parecido a \tilde{O}_j que a O_p , es decir $d(O_j, O_p) \geq d(O_j, \tilde{O}_j)$. Así, si el medoide O_m se reemplaza por O_p , O_j pertenecerá al *cluster* representado por \tilde{O}_j , siendo el coste de este intercambio

$$C_{jmp} = d(O_j, O_{j,2}) - d(O_j, O_m), \quad (3.6)$$

que es siempre no negativo.

- Caso 2. O_j pertenece al *cluster* representado por O_m pero en este caso es más similar a O_p que a \tilde{O}_j , es decir $d(O_j, O_p) < d(O_j, \tilde{O}_j)$. Así el coste del intercambio viene dado por

$$C_{jmp} = d(O_j, O_p) - d(O_j, O_m), \quad (3.7)$$

que puede ser positivo o negativo dependiendo de si O_j es más similar a O_m o a O_p .

- Caso 3. Supongamos que O_j pertenece a otro *cluster* distinto del representado por O_m , siendo \tilde{O}_j el representante de dicho *cluster*. Consideremos que O_j es más parecido a \tilde{O}_j que a O_p . Por ello, si se reemplaza O_m por O_p , O_j seguirá perteneciendo al *cluster* representado por \tilde{O}_j , es decir, el coste del replazo es nulo

$$C_{jmp} = 0. \quad (3.8)$$

- Caso 4. Supongamos finalmente que O_j pertenece al *cluster* representado por \tilde{O}_j y es más similar a O_p que a \tilde{O}_j . Entonces, al reemplazar O_m por O_p , O_j pasará a estar en el *cluster* representado por O_p , siendo el coste

$$C_{jmp} = d(O_j, O_p) - d(O_j, \tilde{O}_j), \quad (3.9)$$

que toma siempre valores negativos.

Combinando estos cuatro casos, el coste total TC_{mp} de reemplazar O_m por O_p viene dado por

$$TC_{mp} = \sum_{j=1}^M C_{jmp}, \quad \text{para } 1 \leq m, p \leq M. \quad (3.10)$$

Por ejemplo, si O_m y O_p son los elementos que diferencian los nodos vecinos S_1 y S_2 , es decir $O_m, O_p \notin S_1 \cap S_2$ pero $O_m \in S_1$ y $O_p \in S_2$, la diferencia entre los costes de los dos nodos vecinos viene dada por la expresión(3.10).

Con esta notación, se puede explicar el algoritmo CLARANS del modo siguiente (véase también el diagrama de flujo en la Figura 3.1):

1. Se establecen los parámetros **numLocal** y **maxVecinos**. Se inicializa $i = 1$ y **minCoste** con un número elevado.
2. Se toma un nodo arbitrario $G_{n,k}$ como **actual**.
3. Se establece $j = 1$.
4. Se considera un vecino arbitrario de **actual**, S . Empleando la ecuación (3.10) se calcula la diferencia entre los costes de los dos nodos.
5. Si S tiene un coste menor, se establece S como **actual** y se repite el proceso desde el paso 3.
6. Si no es menor, se incrementa j en una unidad y si $j \leq \mathbf{maxVecinos}$ se vuelve al paso 4, tomando un nuevo vecino arbitrario.
7. Si $j > \mathbf{maxVecinos}$ se compara el coste de **actual** con **minCoste**, si el coste actual es menor que **minCoste**, se toma el coste de **actual** como **minCoste** y como **mejorNodo** se toma **actual**.
8. Se incrementa el valor de i en una unidad. Si $i > \mathbf{numLocal}$, se detiene el algoritmo. Si no, se regresa al paso 2.

3.3. Método GMEANS

Como se ha podido analizar el método *KMEANS* dependen de los centroides iniciales considerados. Además tanto en *CLARANS* como en *KMEANS* es necesario conocer de antemano en cuantos *clusters* se han de clasificar los datos. Se ha observado que esto puede suponer un problema en el presente caso de estudio dado que lo más habitual es que los compresores funcionen de forma correcta, de manera que se espera que todos los datos sean del mismo tipo. Sin embargo, al preestablecer un número de *clusters* los algoritmos mencionados tratan de clasificar los datos en este número de *clusters* cuando lo correcto sería la existencia de un único grupo.

Es por ello que se plantea el uso de algoritmos que, además de realizar la clasificación, sean capaces de establecer el número de *clusters* necesarios. En un primer momento se planteó el uso del algoritmo *XMEANS* [10]. Este algoritmo ha demostrado gran eficacia a la hora de clasificar los datos, incluso cuando hay muy pocos de un tipo, sin embargo no es capaz de identificar la existencia de un solo *cluster* cuando todos los datos son de un mismo tipo.

Se plantea entonces el uso del algoritmo *GMEANS* [5]. Este algoritmo comienza con un número pequeño de centros y en cada iteración divide en dos los *clusters* e forma que los datos a su alrededor procedan de una distribución Gaussiana. En términos generales, el algoritmo de *GMEANS* implementado en [9] sigue los siguientes pasos:

1. A partir del conjunto de datos totales, se toma un subconjunto $\{C_1, \dots, C_{\tilde{K}}\}$ como el conjunto de centros iniciales. Por lo general se escoge un único representante inicial como la media del conjunto de datos.
2. Se aplica *XMEANS* sobre el conjunto total de datos y como centroides $\{C_1, \dots, C_{\tilde{K}}\}$.
3. Se emplea un test estadístico (véase para más detalle [5]) para determinar si cada *cluster* sigue una distribución Gaussiana con un nivel de confianza α .
4. Si el *cluster* j -ésimo sigue una distribución Gaussiana, se mantiene el centro C_j . Si no se reemplaza C_j por dos centros distintos.
5. Se repite el proceso desde el paso 2 hasta que no es necesario añadir más centros.

Nótese que el algoritmo *GMEANS* emplea el algoritmo *KMEANS* repetidas veces, al menos K (número de *clusters* buscados) veces cuando se encuentran K centros, lo que hace que su tiempo de cálculo sea al menos $O(K)$ veces el tiempo de cálculo del algoritmo *KMEANS*.

Veamos con un poco más de detalle el test estadístico empleado para determinar si los elementos asignados a un centro siguen una distribución Gaussiana. El planteamiento del test es el siguiente:

- H_0 : Los datos asignados al *cluster* siguen una distribución Gaussiana (hipótesis nula).
- H_1 : Los datos asignados al *cluster* no siguen una distribución Gaussiana (hipótesis alternativa).

Si no rechazamos la hipótesis nula H_0 , consideramos que ese centro es suficiente para representar los datos que se le asignan y no se divide el *cluster* actual en dos. Si rechazamos H_0 , entonces el *cluster* debe dividirse en dos.

El test empleado se basa en el estadístico de Anderson-Darling (véase [7] para más detalles). Dada una lista de valores escalares y_i con $i = 1, \dots, n$ que han sido tipificados, denotemos por $y_{(i)}$ al i -ésimo valor ordenado. Sea $z_i = F(y_{(i)})$, donde F es la función de distribución normal de media nula y varianza unidad, el estadístico en cuestión es

$$A^2(Z) = -\frac{1}{n} \sum_{i=1}^n (2i-1) [\log z_i + \log (1 - z_{n+1-i})] - n. \quad (3.11)$$

Dado que μ y σ se estiman de los datos, se realiza la siguiente corrección

$$A_*^2(Z) = A^2(Z)(1 + 4/n - 25/n^2). \quad (3.12)$$

De este modo, dado un subconjunto de datos (candidato a ser un *cluster*) asignados al centro C_k , el test se lleva a cabo del siguiente modo

1. Se elige un nivel significativo α para el test.
2. Se toman dos centros, llamados hijos del centro c .
3. Se aplica el algoritmo *KMEANS* para estos dos centros hijos y el *cluster*. Se denotan por C_{k1} y C_{k2} a los dos centros seleccionados por *KMEANS*.

4. Se calcula la diferencia entre ambos centros $D = C_{k1} - C_{k2}$ y se proyecta cada elemento del *cluster* sobre D

$$O'_i = \frac{\text{Proyección}(O_i, D)}{\|D\|^2}. \quad (3.13)$$

Así, cada uno de los elementos O'_i es la representación unidimensional de los datos proyectados sobre D . El conjunto de proyecciones asociados a D y al *cluster* k -ésimo se tipifica de manera que tenga media 0 y varianza 1.

5. Sea $z_i = F(O'_{(i)})$. Si $A_*^2(Z)$ está en el rango de valores que no son críticos con un nivel de confianza α , entonces no rechazamos H_0 , manteniendo el centro original C_k y descartando los centros $\{C_{k1}, C_{k2}\}$. En caso contrario, se rechaza H_0 y se sustituye el centro original C_k por $\{C_{k1}, C_{k2}\}$.

Como se indica en el paso 2, es necesario inicializar dos centros hijos C_{k1} y C_{k2} , esta elección suele hacerse de dos formas distintas en este tipo de algoritmos. Ambos enfoques consideran una inicialización de la forma $C_k \pm A$, donde C_k es el centro y A varía según el enfoque escogido. En el primer enfoque se escoge un elemento A aleatorio, de modo que $\|A\|$ sea pequeño en comparación con la variabilidad de los datos. El segundo método halla la componente principal de los datos S , con autovalor λ , y se fija $A = S\sqrt{2\lambda/\pi}$.

Capítulo 4

Clasificación de datos manufacturados

En este capítulo se presentará cómo se implementan los métodos explicados en el capítulo anterior y se realizarán algunas pruebas para verificar la precisión de los distintos algoritmos.

Para ello se generarán 100 curvas que son soluciones del sistema de ecuaciones diferenciales descritas en el Capítulo 2. En particular, cada una de las curvas se ha discretizado usando 100 valores equiespaciados a lo largo de un único periodo. Para generar estas curvas manufacturadas se han utilizado distintos valores de los coeficientes de rigidez. Nos interesa analizar, por tanto, si los algoritmos de clasificación son capaces de distinguir entre las curvas solución que proceden de coeficientes distintos.

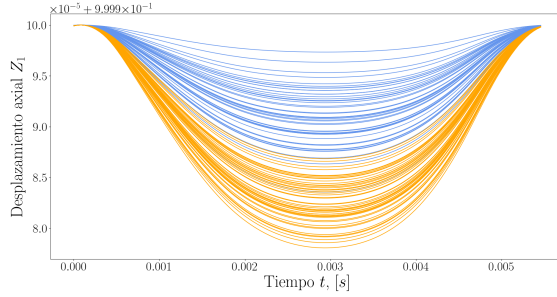
4.1. Clasificación con clusters con el mismo número de elementos

En primer lugar se plantea si los algoritmos son capaces de clasificar las curvas de manera que la mitad sean de un tipo y la otra mitad de otro, es decir, son curvas que se han calculado con coeficientes de rigidez que se generan como $1 \cdot 10^3 + 300\varepsilon$ con $\varepsilon \sim N(0, 1)$, esto es, ε representa un error de media nula y varianza unidad y de distribución normal. Las otras 50 curvas se han calculado con coeficientes de rigidez que se generan como $2 \cdot 10^3 + 300\varepsilon$ con $\varepsilon \sim N(0, 1)$. Contamos con cuatro coeficientes de rigidez que podemos hacer variar simultáneamente pero para hacer más sencillo el análisis de los

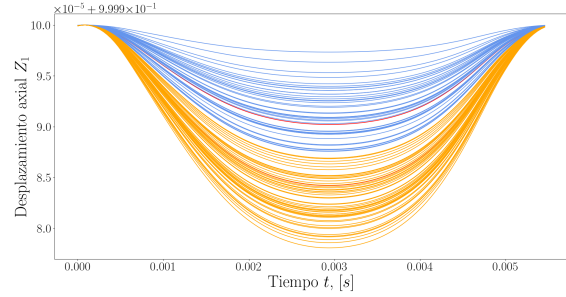
resultados, tan sólo hacemos variar un par fijando el resto.

4.1.1. Clasificación según los coeficientes de rigidez axial de los rotores

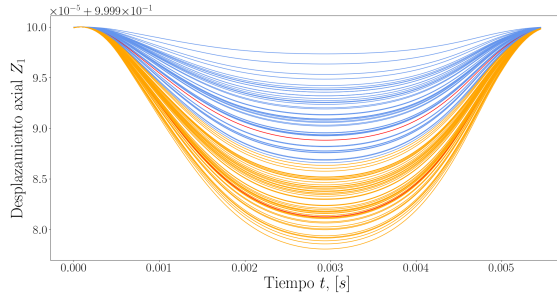
En primer lugar se realiza la clasificación considerando que los coeficientes relacionados con la lubricación de aceite de los rotores son $k_{0,a} = 10^3$ y $k_{0,t} = 10^3$, y $k_3 = 10^4$, mientras que los coeficientes de rigidez asociados al movimiento axial de los rotores, k_1 y k_2 toman los valores señalados anteriormente (esto es, sus 50 primeros valores dados por $10^3 + 300\varepsilon$ y los 50 restantes por $2 \cdot 10^3 + 300\varepsilon$)



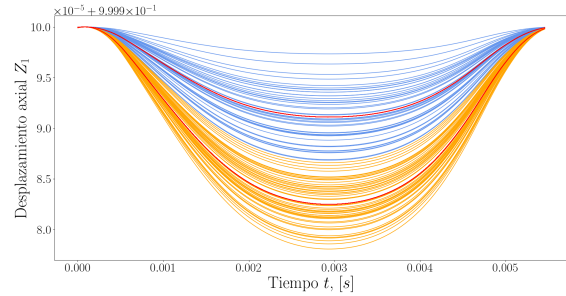
(a) Clasificación real de los desplazamientos Z_1



(b) Clasificación con *CLARANS*



(c) Clasificación con *KMEANS*



(d) Clasificación con *GMEANS*

Figura 4.1: Clasificación de los desplazamientos Z_1 obtenidos considerando $k_{0,a} = 10^3$, $k_{0,t} = 10^3$ y $k_3 = 10^4$ constantes y variando los coeficientes de rigidez axial k_1 y k_2 .

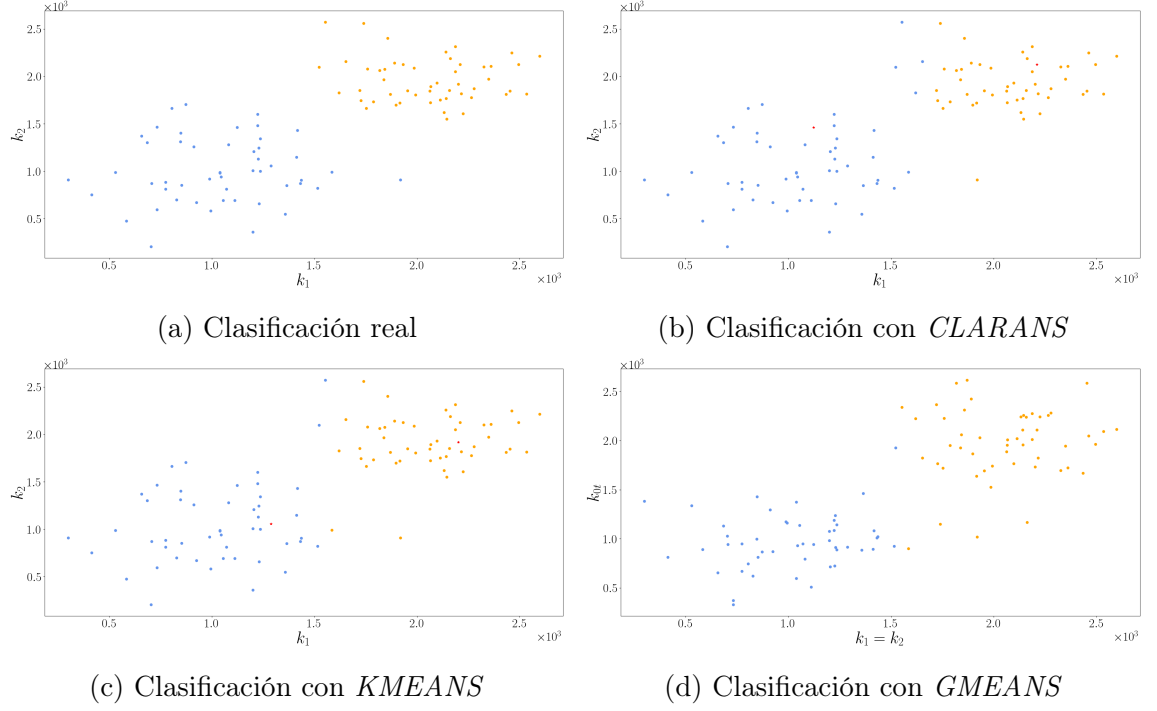


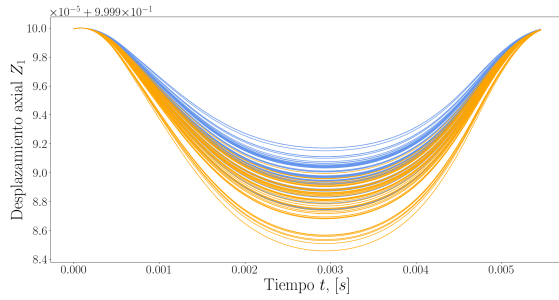
Figura 4.2: Clasificación de los parámetros k_1 y k_2 considerando que sus 50 primeros valores vienen dados por $10^3 + 300\varepsilon$ y los restantes por $2 \cdot 10^3 + 300\varepsilon$.

En las Figuras 4.1 y 4.2 se puede observar que todos los métodos realizan la clasificación de manera precisa. En la Figura 4.1 aparecen señalados en color rojo las curvas que los métodos consideran el centro (medoide o centroide) de cada *cluster*.

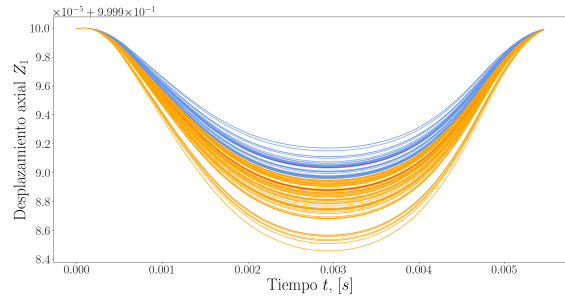
4.1.2. Clasificación según los coeficientes de rigidez de la lubricación del aceite

En este caso los coeficientes de rigidez de la lubricación del aceite, $k_{0,a}$ y $k_{0,t}$, toman los valores señalados anteriormente (esto es, sus 50 primeros valores dados por $10^3 + 300\varepsilon$ y los 50 restantes por $2 \cdot 10^3 + 300\varepsilon$), mientras que los coeficientes de rigidez axial, $k_1 = 10^3$ y $k_2 = 10^3$, y $k_3 = 10^4$ se considerarán constantes. Como se puede ver en las Figuras 4.3 y 4.4 estas clasificaciones no son tan precisas como las de la sección anterior, esto se distingue sobre todo en la Figura 4.4, pues claramente hay muchos puntos

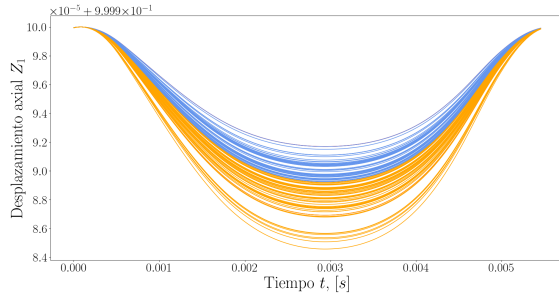
mal clasificados. Cabe destacar que en este caso *GMEANS* determina que las curvas se clasifican mejor en tres *clusters* que en dos.



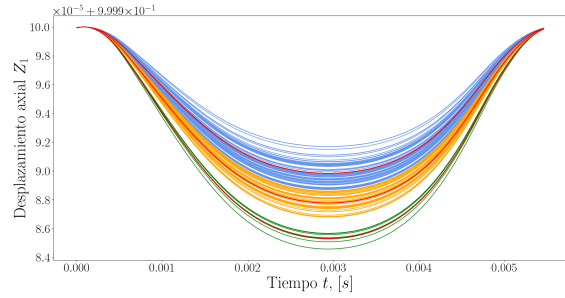
(a) Clasificación real de los desplazamientos Z_1



(b) Clasificación con *CLARANS*



(c) Clasificación con *KMEANS*



(d) Clasificación con *GMEANS*

Figura 4.3: Clasificación de los desplazamientos Z_1 obtenidos considerando $k_1 = 10^3$, $k_1 = 10^3$ y $k_3 = 10^4$ constantes y variando los coeficientes $k_{0,a}$ y $k_{0,t}$.

4.1. CLASIFICACIÓN CON CLUSTERS CON EL MISMO NÚMERO DE ELEMENTOS 31

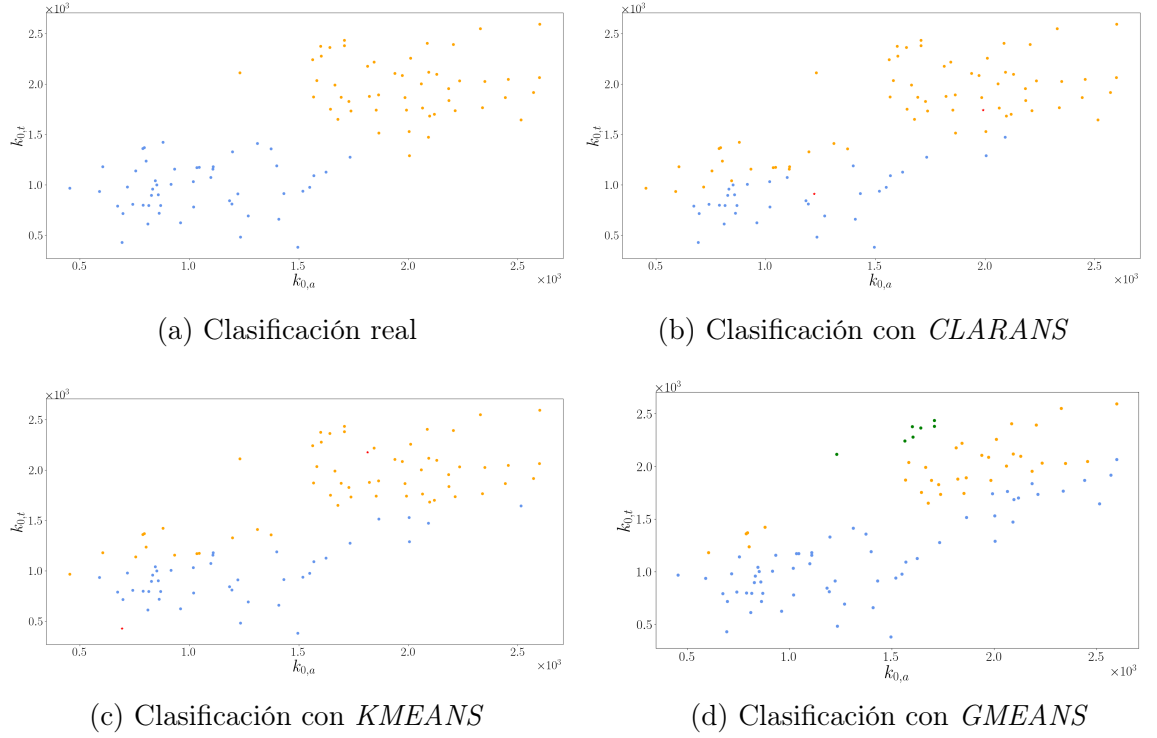
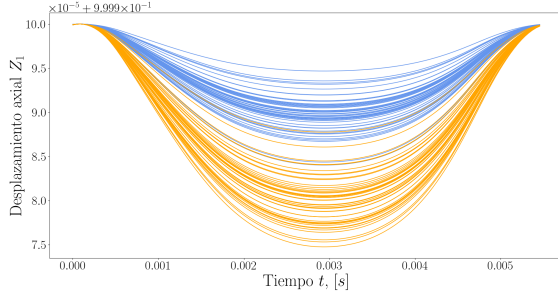


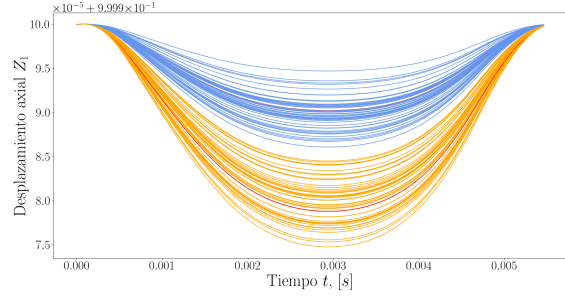
Figura 4.4: Clasificación de los parámetros $k_{0,a}$ y $k_{0,t}$ considerando que sus 50 primeros valores vienen dados por $10^3 + 300\varepsilon$ y los restantes por $2 \cdot 10^3 + 300\varepsilon$.

4.1.3. Clasificación según los coeficientes de rigidez axial de los rotores y de la lubricación del aceite

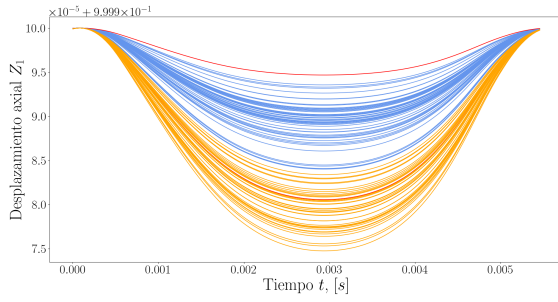
Se considerarán ahora $k_1 = k_2$, $k_{0,a}$ y $k_{0,t}$ con sus primeros 50 elementos de la forma $10^3 + 300\varepsilon$ y los 50 siguientes de la forma $2 \cdot 10^3 + 300\varepsilon$. $k_3 = 10^4$ se toma constante. De nuevo en las Figuras 4.5 y 4.6 se observa que la clasificación es correcta.



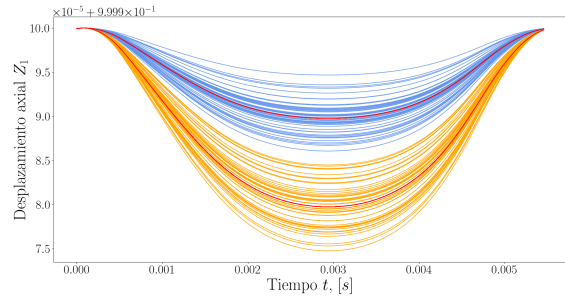
(a) Clasificación real de los desplazamientos Z_1



(b) Clasificación con *CLARANS*

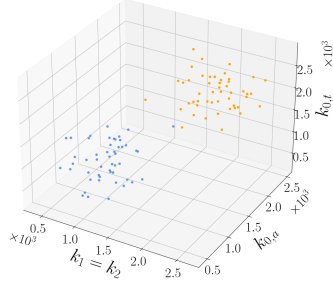


(c) Clasificación con *KMEANS*

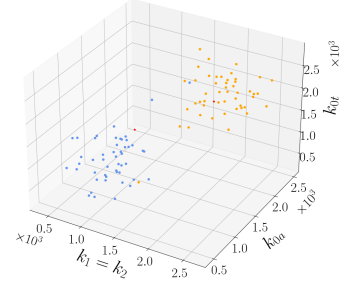


(d) Clasificación con *GMEANS*

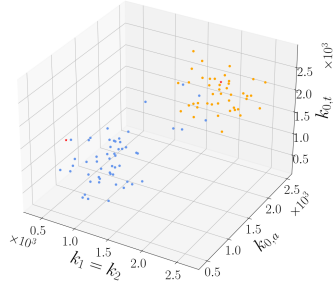
Figura 4.5: Clasificación de los desplazamientos Z_1 obtenidos considerando $k_3 = 10^4$ constante y variando los coeficientes $k_1 = k_2$, $k_{0,a}$ y $k_{0,t}$.



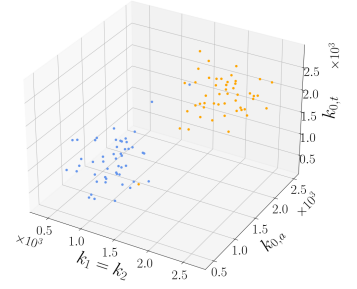
(a) Clasificación real



(b) Clasificación con *CLARANS*



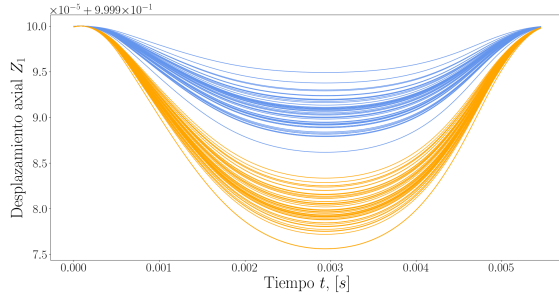
(c) Clasificación con *KMEANS*



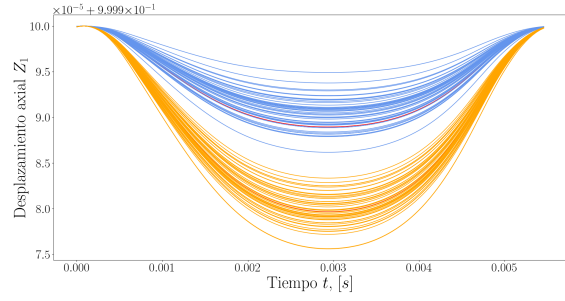
(d) Clasificación con *GMEANS*

Figura 4.6: Clasificación de los parámetros $k_1 = k_2$, $k_{0,a}$ y $k_{0,t}$ considerando que sus 50 primeros valores vienen dados por $10^3 + 300\varepsilon$ y los restantes por $2 \cdot 10^3 + 300\varepsilon$.

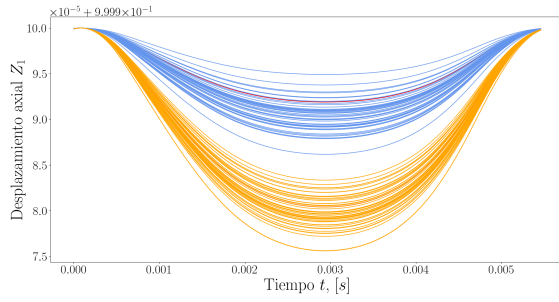
De modo análogo tomando los 50 primeros elementos de k_1 , k_2 y $k_{0,a} = k_{0,t}$ de la forma $10^3 + 300\varepsilon$, los 50 siguientes de la forma $2 \cdot 10^3 + 300\varepsilon$ y $k_3 = 10^4$ constante se obtienen las Figuras 4.7 y 4.8, donde se observa que la clasificación es correcta.



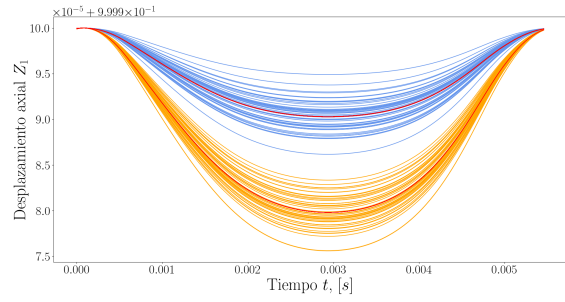
(a) Clasificación real de los desplazamientos Z_1



(b) Clasificación con *CLARANS*

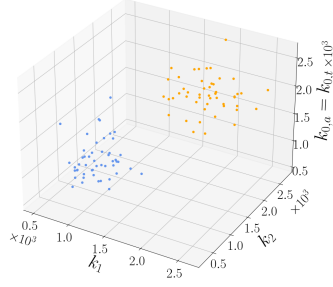


(c) Clasificación con *KMEANS*

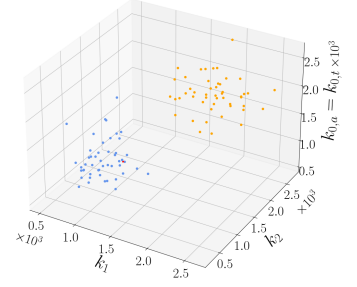


(d) Clasificación con *GMEANS*

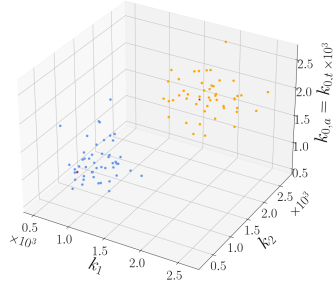
Figura 4.7: Clasificación de los desplazamientos Z_1 obtenidos considerando $k_3 = 10^4$ constante y variando los coeficientes $k_1 = k_2$, $k_{0,a}$ y $k_{0,t}$.



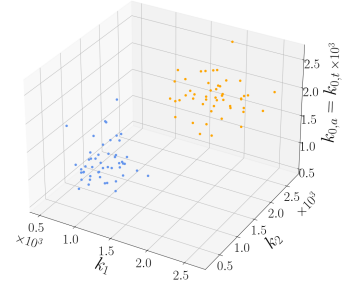
(a) Clasificación real



(b) Clasificación con *CLARANS*



(c) Clasificación con *KMEANS*

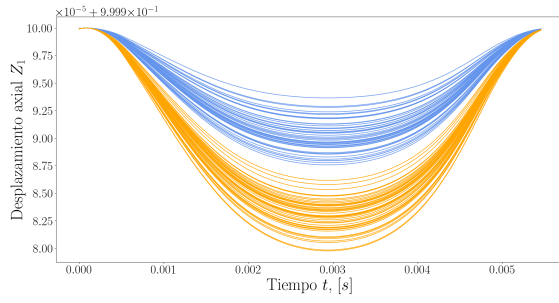


(d) Clasificación con *GMEANS*

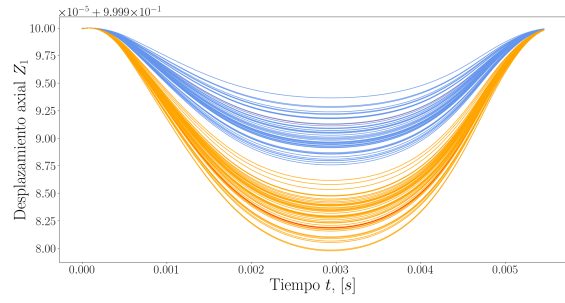
Figura 4.8: Clasificación de los parámetros $k_1, k_2, k_{0,a} = k_{0,t}$ considerando que sus 50 primeros valores vienen dados por $10^3 + 300\varepsilon$ y los restantes por $2 \cdot 10^3 + 300\varepsilon$.

4.1.4. Clasificación según los coeficientes de rigidez axial y el coeficiente de rigidez torsional conjunta

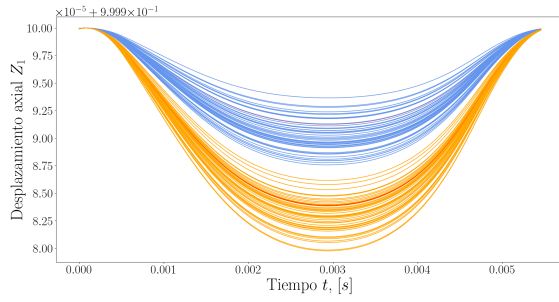
En este caso se tomarán k_1, k_2 y k_3 con sus primeros 50 elementos de la forma $10^3 + 300\varepsilon$ y los 50 restantes de la forma $2 \cdot 10^3 + 300\varepsilon$, mientras que $k_{0,a} = k_{0,t} = 10^3$ toman un valor constante.



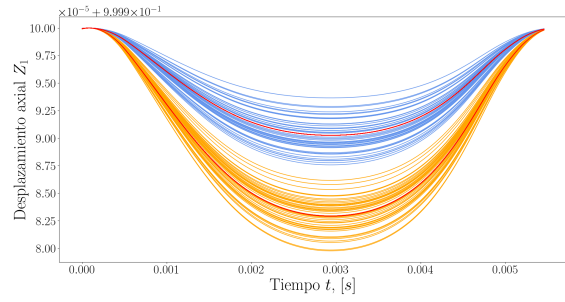
(a) Clasificación real de los desplazamientos Z_1



(b) Clasificación con *CLARANS*

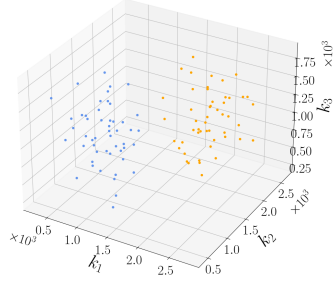


(c) Clasificación con *KMEANS*



(d) Clasificación con *GMEANS*

Figura 4.9: Clasificación de los desplazamientos Z_1 obtenidos considerando $k_{0,a} = k_{0,t} = 10^3$ constante y variando los coeficientes $k_1 = k_2$, k_2 y k_3 .



(a) Clasificación real

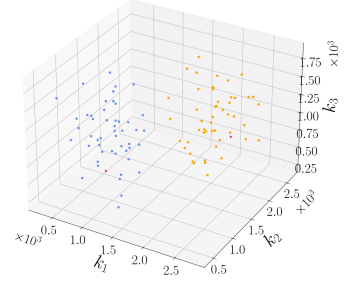
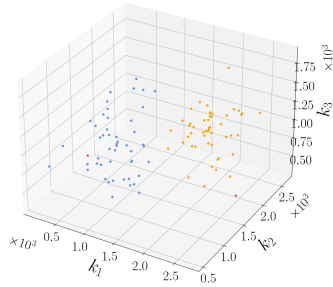
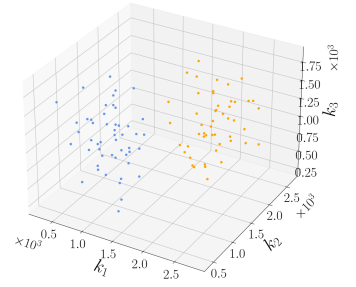
(b) Clasificación con *CLARANS*(c) Clasificación con *KMEANS*(d) Clasificación con *GMEANS*

Figura 4.10: Clasificación de los parámetros k_1 , k_2 , k_3 considerando que sus 50 primeros valores vienen dados por $10^3 + 300\varepsilon$ y los restantes por $2 \cdot 10^3 + 300\varepsilon$.

En las Figuras 4.9 y 4.10 se aprecia de nuevo una buena clasificación por parte de todos los métodos.

4.2. Clasificación con un cluster con muchos menos elementos

Como se ha mencionado, es necesario que el algoritmo sea capaz de clasificar correctamente en el caso de que casi todos los datos sean de un tipo y solo unos pocos sean de otro tipo. Para comprobar qué algoritmo tiene un mejor comportamiento en estos casos se trata de clasificar cien curvas de desplazamientos, siendo cinco de un tipo y el resto de otro. Para hacer este estudio se han generado las curvas con los siguientes datos: las primeras 95 son curvas que se han calculado con coeficientes de rigidez axial k_1 y k_2 que

se generan como $1 \cdot 10^3 + 300\varepsilon$ con $\varepsilon \sim N(0, 1)$, esto es, ε representa un error de media nula y varianza unidad y de distribución normal. Las otras 5 curvas se han calculado con coeficientes de rigidez axial k_1 y k_2 que se generan como $2 \cdot 10^3 + 300\varepsilon$ con $\varepsilon \sim N(0, 1)$

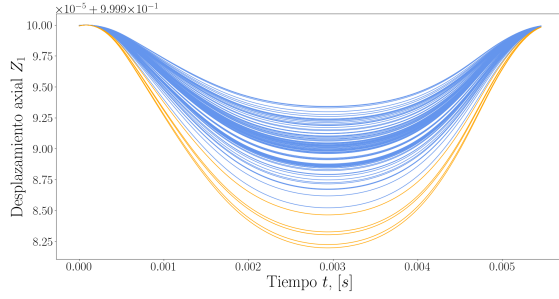
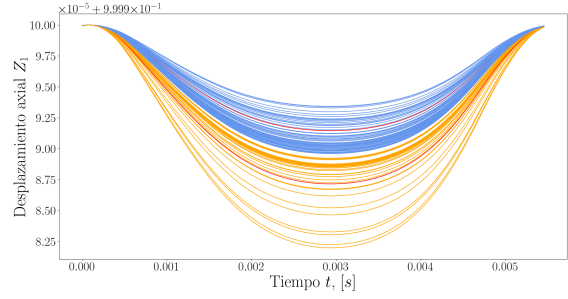
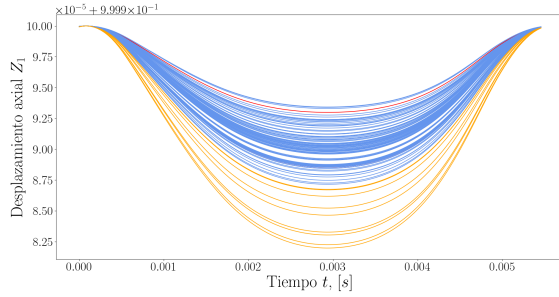
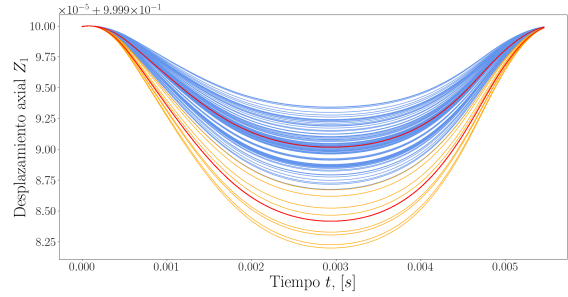
(a) Clasificación real de los desplazamientos Z_1 (b) Clasificación con *CLARANS*(c) Clasificación con *KMEANS*(d) Clasificación con *GMEANS*

Figura 4.11: Clasificación de los desplazamientos Z_1 obtenidos considerando constantes $k_{0,a} = k_{0,t} = 10^3$ y $k_3 = 10^4$ y variando los coeficientes k_1 y k_2

Como se puede ver en las Figuras 4.11 y 4.12, *CLARANS* no realiza una clasificación satisfactoria, mientras que *KMEANS* y *GMEANS* sí lo hacen.

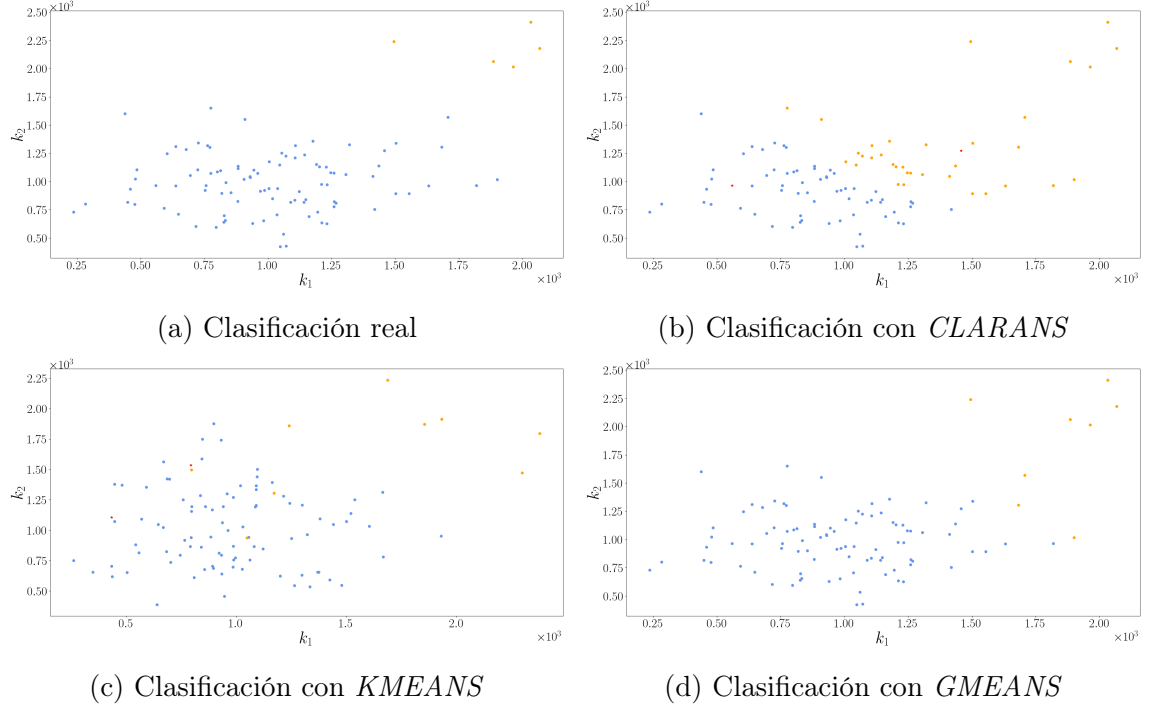
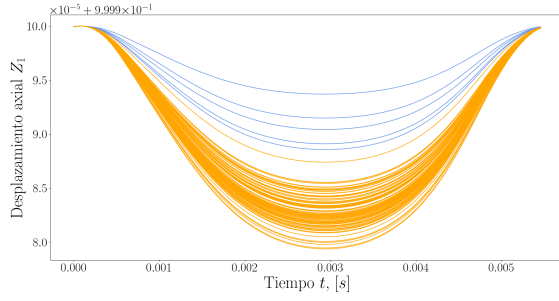


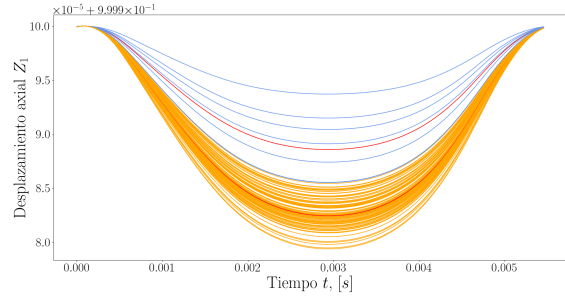
Figura 4.12: Clasificación de los parámetros k_1 y k_2 considerando que sus 95 primeros valores vienen dados por $10^3 + 300\varepsilon$ y los 5 restantes por $2 \cdot 10^3 + 300\varepsilon$.

Se ha realizado también la clasificación considerando que las primeras 5 son curvas que se han calculado con coeficientes de rigidez generados como $1 \cdot 10^3 + 300\varepsilon$ con $\varepsilon \sim N(0, 1)$, esto es, ε representa un error de media nula y varianza unidad y de distribución normal. Las otras 95 curvas se han calculado con coeficientes de rigidez que se generan como $2 \cdot 10^3 + 300\varepsilon$ con $\varepsilon \sim N(0, 1)$. Nótese que al contrario que en el caso anterior, el *cluster* con mayor número de elementos es el que toma los valores elevados.

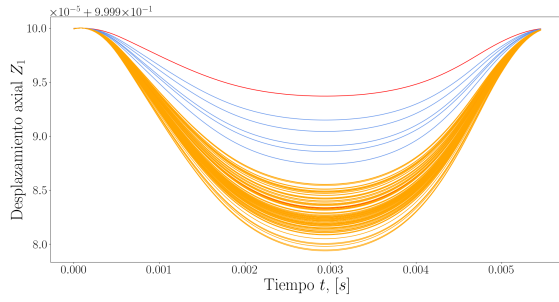
En las Figuras 4.13 y 4.14 se observa que en este caso todos los algoritmos clasifican satisfactoriamente., incluso *CLARANS*.



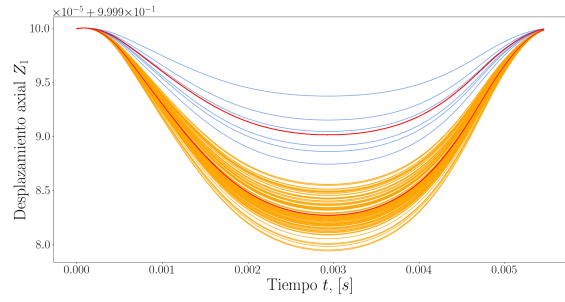
(a) Clasificación real de los desplazamientos Z_1



(b) Clasificación con *CLARANS*



(c) Clasificación con *KMEANS*



(d) Clasificación con *GMEANS*

Figura 4.13: Clasificación de los desplazamientos Z_1 obtenidos considerando constantes $k_{0,a} = k_{0,t} = 10^3$ y $k_3 = 10^4$ y variando los coeficientes k_1 y k_2

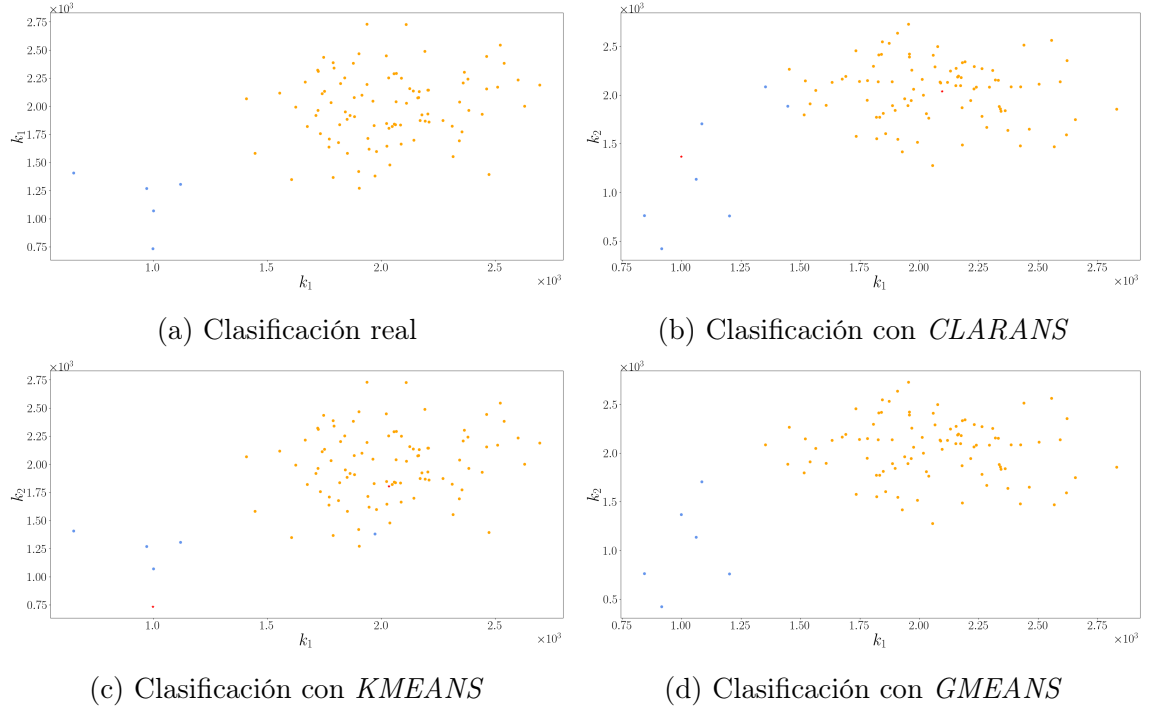


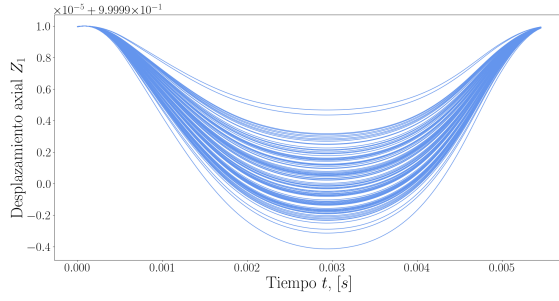
Figura 4.14: Clasificación de los parámetros k_1 y k_2 considerando que sus 95 primeros valores vienen dados por $10^3 + 300\epsilon$ y los 5 restantes por $2 \cdot 10^3 + 300\epsilon$.

4.3. Clasificación con un único cluster

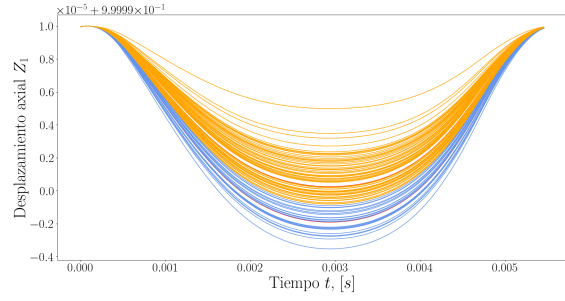
Puesto que los datos reales de los que se disponen por parte de la empresa son de un compresor en condiciones de funcionamiento correctas, es necesario que el algoritmo de clasificación sea capaz de distinguir este comportamiento.

Esto, sin embargo es imposible con los algoritmos *CLARANS* y *KMEANS*, ya que es necesario introducir *a priori* el número de *clusters* necesarios. Si establecemos que estos algoritmos deben identificar un único *cluster* no serán capaces de detectar errores en ningún caso, pero si establecemos que deben clasificar los datos en dos *clusters* siempre tenderán a asignar alguna curva a cada uno de los *clusters*, como se puede ver en las Figuras 4.15b, 4.16b y 4.15c, 4.16c.

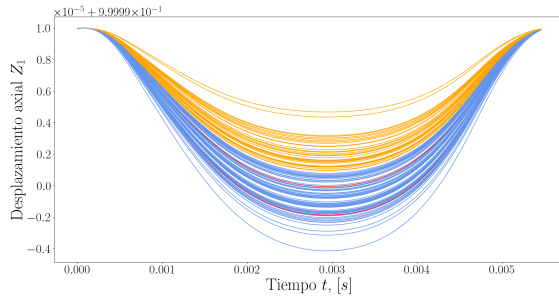
Por ello se plantea el uso del algoritmo *GMEANS*, ya que este parte de un solo *cluster* y, si es necesario, añade más. En efecto, en las Figuras 4.15d y 4.16d se puede ver que la clasificación se realiza en un único *cluster*.



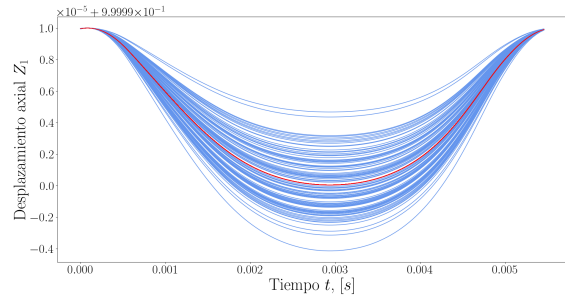
(a) Clasificación real de los desplazamientos Z_1 .



(b) Clasificación con *CLARANS*



(c) Clasificación con *KMEANS*



(d) Clasificación con *GMEANS*

Figura 4.15: Clasificación de los desplazamientos Z_1 obtenidos considerando constantes $k_{0,a} = k_{0,t} = 10^3$ y $k_3 = 10^4$ y variando los coeficientes k_1 y k_2

4.4. CLASIFICACIÓN USANDO DATOS MANUFACTURADOS CON RUIDO 43

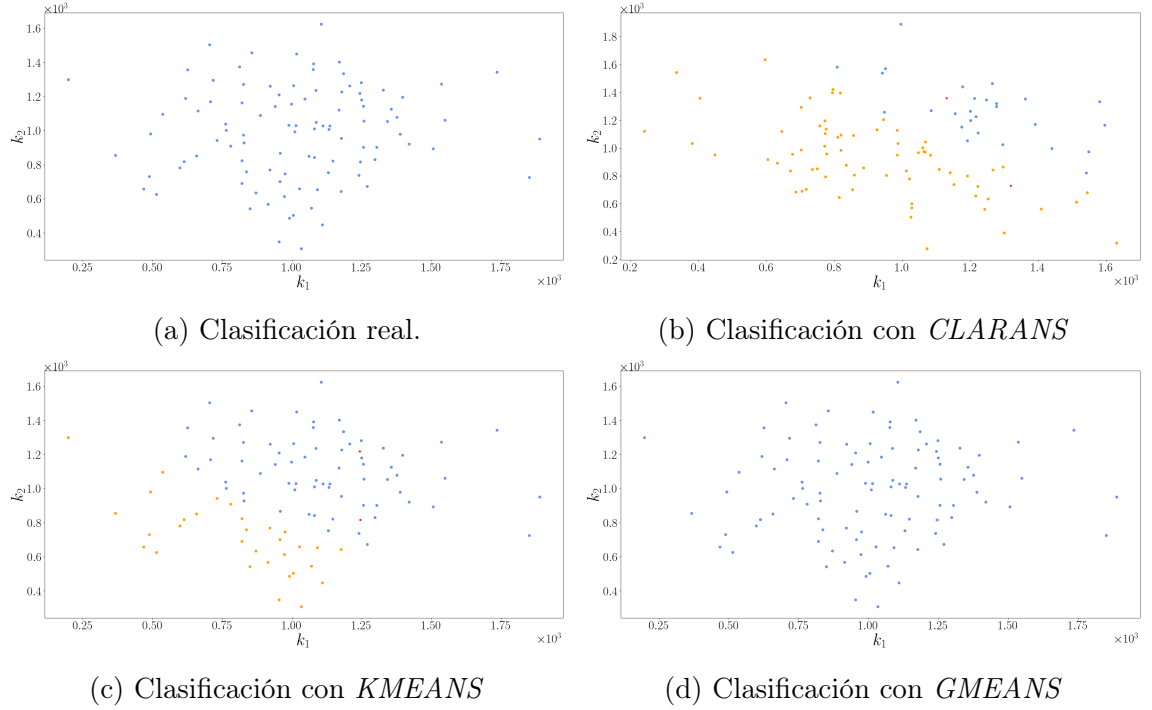
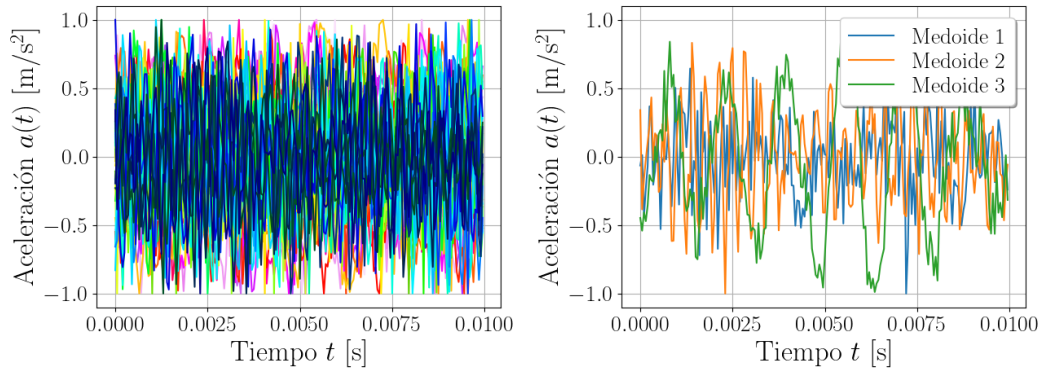


Figura 4.16: Clasificación de los parámetros k_1 y k_2 considerando que sus 100 valores vienen dados por $10^3 + 300\varepsilon$.

4.4. Clasificación usando datos manufacturados con ruido

En los datos manufacturados previamente presentados las curvas a clasificar eran muy suaves, ya que se generaban como soluciones de un sistema de ecuaciones diferenciales ordinarias, lineales, con coeficientes constantes y donde sus segundos miembros también eran suaves.

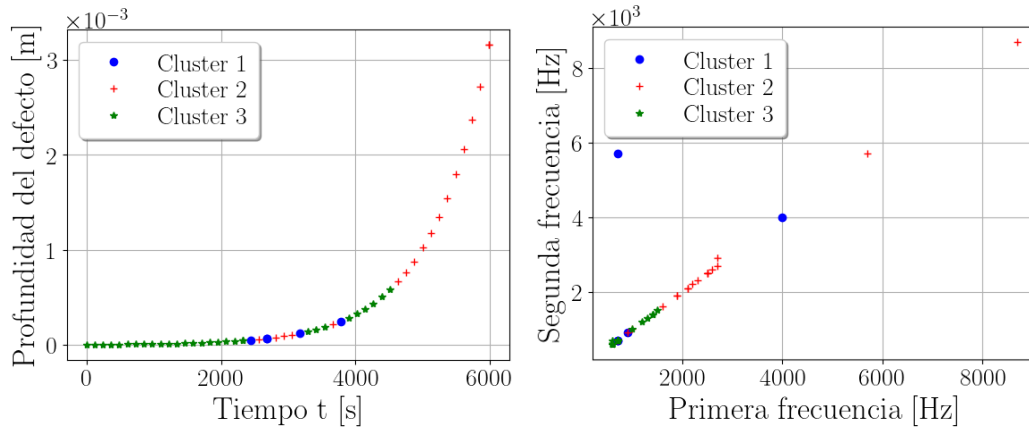
Sin embargo, como se verá en el próximo capítulo, los datos reales presentan un alto nivel de oscilaciones. Para tratar este tipo de datos, en esta sección se trabajará con datos manufacturados de vibraciones de un rodamiento de bolas que se obtienen mediante simulación numérica (véase [2] para más detalles). Se dispone de datos de vibraciones para rodamientos en distinto estado de desgaste (de entre $3\mu m$ y $3mm$). En la Figura 4.17a se representan los datos de vibraciones.



(a) Señales de vibraciones contenidas en la base de datos. (b) Señales de vibraciones asociadas a los medoides.

Figura 4.17: Señales de vibraciones de un rodamiento de bolas generado por simulación numérica (véase [2]).

Al clasificar estas señales en tres *clusters* mediante *CLARANS* se obtienen tres medoides, asociados cada uno a una señal de vibración, que se pueden interpretar como representantes de cada *cluster*. Estas tres curvas aparecen representadas en la Figura 4.17b.



(a) Clasificación según del desgaste de los rodamientos. (b) Clasificación según las frecuencias de las vibraciones.

De cara a poder decidir si el algoritmo es capaz de distinguir un buen o mal funcionamiento del rodamiento se intenta relacionar un mejor funcionamiento cuanto menor sea el desgaste. En la Figura 4.18a se puede ver la evolución del

desgaste a lo largo del tiempo y se representa en distintos colores los distintos *clusters*. En la Figura 4.18b se aprecia como las frecuencias más elevadas (marcadas con cruces en rojo) se corresponden con un mayor desgaste de los rodamientos.

El algoritmo distingue correctamente las señales de vibración que se corresponden tanto con desgastes muy pequeños (mejor funcionamiento), como con desgastes grandes (peor funcionamiento). Este ejemplo ilustra el hecho general de que las frecuencias de vibración más elevadas están asociadas a un mayor desgaste del aparato o a un posible mal funcionamiento.

Capítulo 5

Clasificación de datos reales

En la última etapa de este proyecto se pretenden trasladar las experiencias realizadas con datos experimentales a los recogidos mediante vibrómetros situados en diferentes lugares del compresor. Se disponen de datos de velocidades de vibraciones de cuatro vibrómetros situados en la parte delantera y trasera del motor y del compresor, mostrados en la Figura 5.1.

Las medidas son recogidas aproximadamente cada dos minutos, lo que supone contemplar un enfoque distinto que el empleado anteriormente. En efecto, la resolución del sistema de ecuaciones (2.1) nos permitía obtener datos de vibraciones en una revolución del compresor, es decir, en tiempos del orden de la milésima de segundo. Claramente, los datos reales están en una escala de tiempo totalmente distinta y no podemos clasificar las vibraciones tomando como referencia cada una de las revoluciones del compresor. *COFRICO* ha proporcionado toda la información mostrada en las Figuras 5.2 y 5.3, relacionada con el estado del compresor, de esta manera es posible asociar los regímenes de funcionamiento del compresor con sus vibraciones. Como se puede ver, se tienen datos de potencia eléctrica, frecuencia de trabajo, intensidad, posición de la corredera, presión de aspiración y descarga y temperatura de aspiración y descarga en distintos intervalos de tiempo.

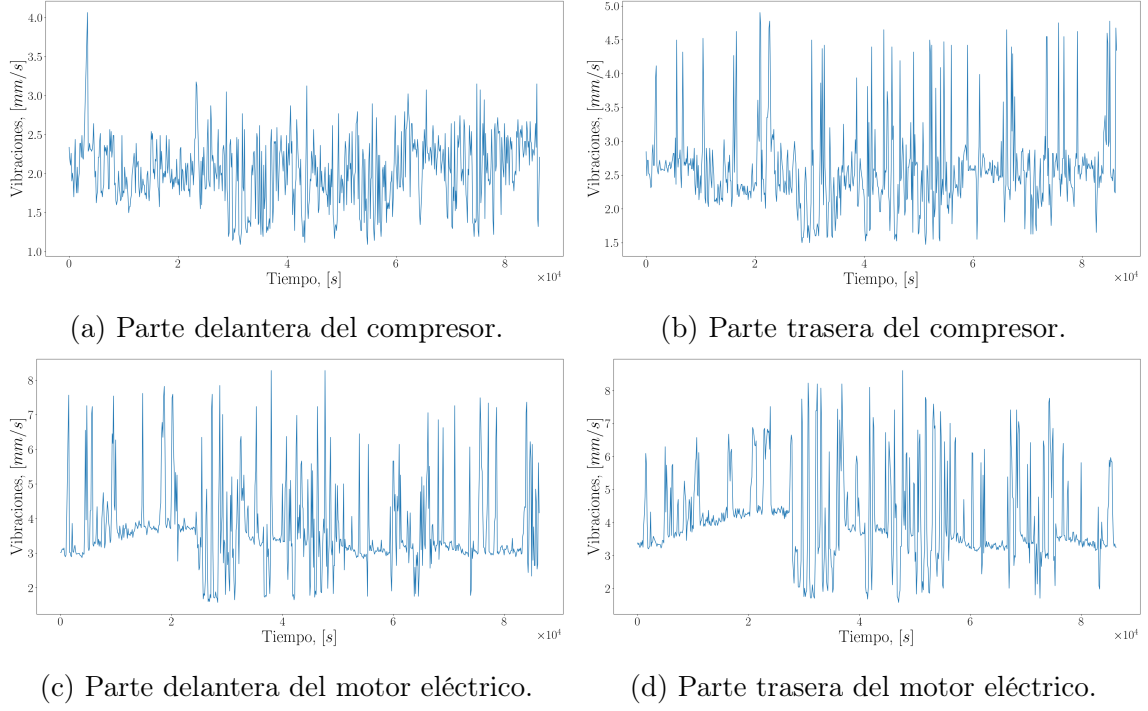


Figura 5.1: Velocidades captadas por los distintos vibrómetros a lo largo de un día.

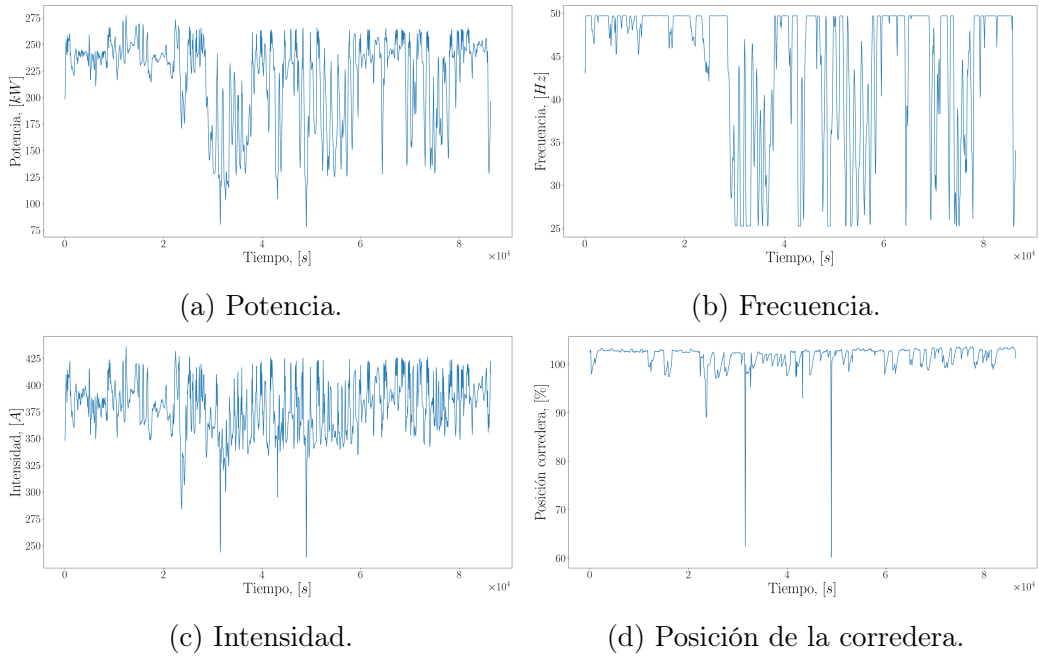


Figura 5.2: Regímenes de potencia, frecuencia, intensidad y posición de corredera a lo largo de un día.

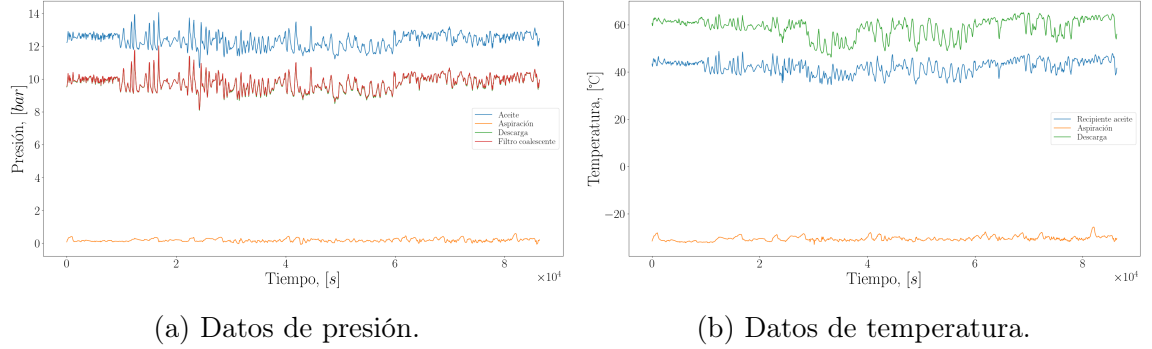


Figura 5.3: Presión y temperatura en distintos procesos del compresor a lo largo de un día.

Se han considerado los tiempos donde la posición de la corredera toma un valor casi constante, para evitar los errores de clasificación derivados de que el compresor se haya apagado. El parámetro de la posición de la corredera toma valores relativos entre 0 % y 100 %, siendo cero cuando el compresor está apagado y aproximadamente 100 % cuando está encendido. Nos interesa clasificar únicamente los valores en los que el compresor está en funcionamiento, por lo que solamente se considerarán valores en los que el parámetro de la posición de la corredera sea superior al 90 %. Puesto que las series de tiempo proporcionadas por la empresa incluyen arranques y paradas del compresor, es necesario extraer únicamente tramos donde las condiciones de trabajo del compresor sean constantes. Obtenemos curvas de velocidades con cinco puntos cada una (se corresponden con las medidas tomadas a lo largo de 10 minutos) asociadas a un funcionamiento estable del compresor. De esta manera las curvas podrán ser clasificadas con los métodos vistos anteriormente.

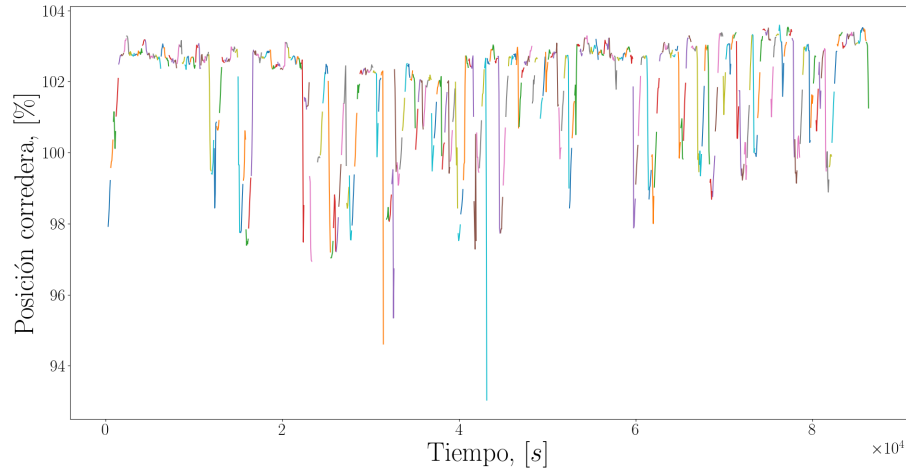
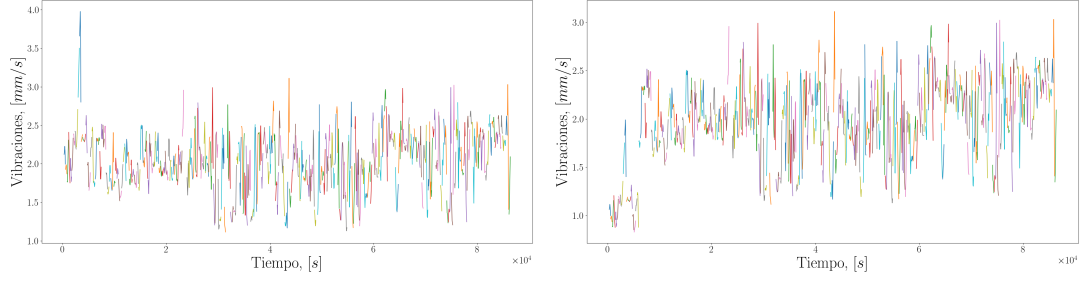
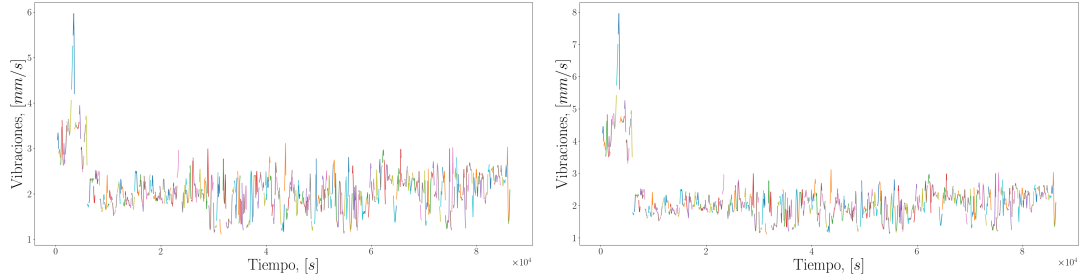


Figura 5.4: Posición de la corredera y vibraciones en los datos experimentales a lo largo de un día.

En la Figura 5.4, la posición de la corredera se representa gráficamente. Están representados en colores diferentes los distintos fragmentos en donde la gráfica toma valores superiores al 90 %. La Figura 5.5a muestra los datos de velocidad captados por el vibrómetro ubicado en la parte delantera del compresor y se representan en colores los fragmentos que se corresponden con tiempos en los que la posición de la corredera es superior al 90 %.



(a) Vibraciones reales. Sin introducir error. (b) Vibraciones introduciendo un error de factor 0.5 las primeras 20 curvas .



(c) Vibraciones introduciendo un error de factor 1.5 las primeras 20 curvas. (d) Vibraciones introduciendo un error de factor 2 las primeras 20 curvas.

Figura 5.5: Detalle de los segmentos de las gráficas empleados en las pruebas. Cada uno de los distintos colores representa una curva de 5 puntos en la que el compresor presenta condiciones de trabajo constantes.

Tal y como se ha mencionado en repetidas ocasiones, el compresor en el que se han realizado las medidas experimentales no presenta datos de un funcionamiento incorrecto, con lo que no se puede comprobar que el algoritmo pueda clasificarlos correctamente. Para poder estimar qué error es capaz de detectar cada uno de los métodos de clasificación, la empresa ha sugerido que un subconjunto de curvas se multiplique por un factor con el objetivo de relacionar estas curvas con datos de malfuncionamiento o de ruido. En lo que sigue, se estudiará con qué valores de este factor es capaz el clasificador de distinguir las curvas asociadas a un funcionamiento correcto o incorrecto. En la Figura 5.5 se muestran algunas de las pruebas realizadas, que se obtienen multiplicado veinte curvas por distintos factores.

Representando estas curvas en un mismo intervalo de tiempo, recuérdese que será de diez minutos, y aplicando los algoritmos de clasificación explicados en capítulos anteriores (estos son: *CLARANS*, *KMEANS* y *GMEANS*)

se puede ver que los algoritmos clasifican satisfactoriamente los casos en los que el factor de error es menor o igual a 0.5 o bien es mayor a 1.5.

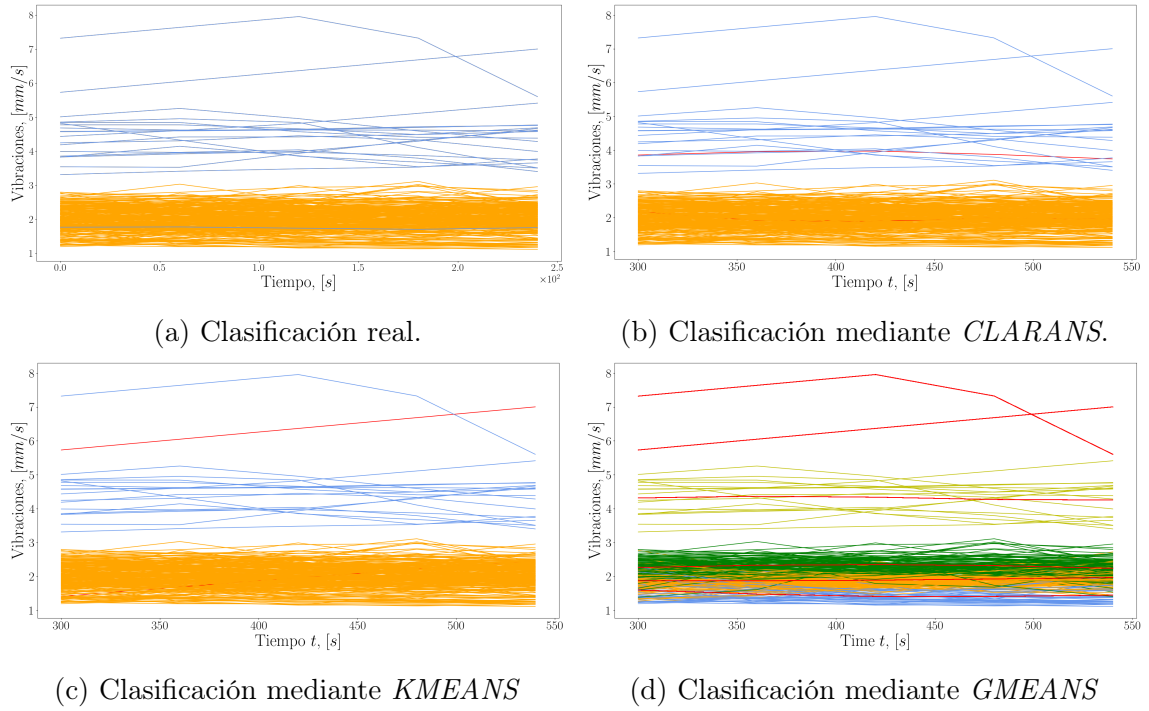


Figura 5.6: Clasificación de las curvas de velocidad introduciendo un factor de error de 2 sobre las primeras 20 curvas.

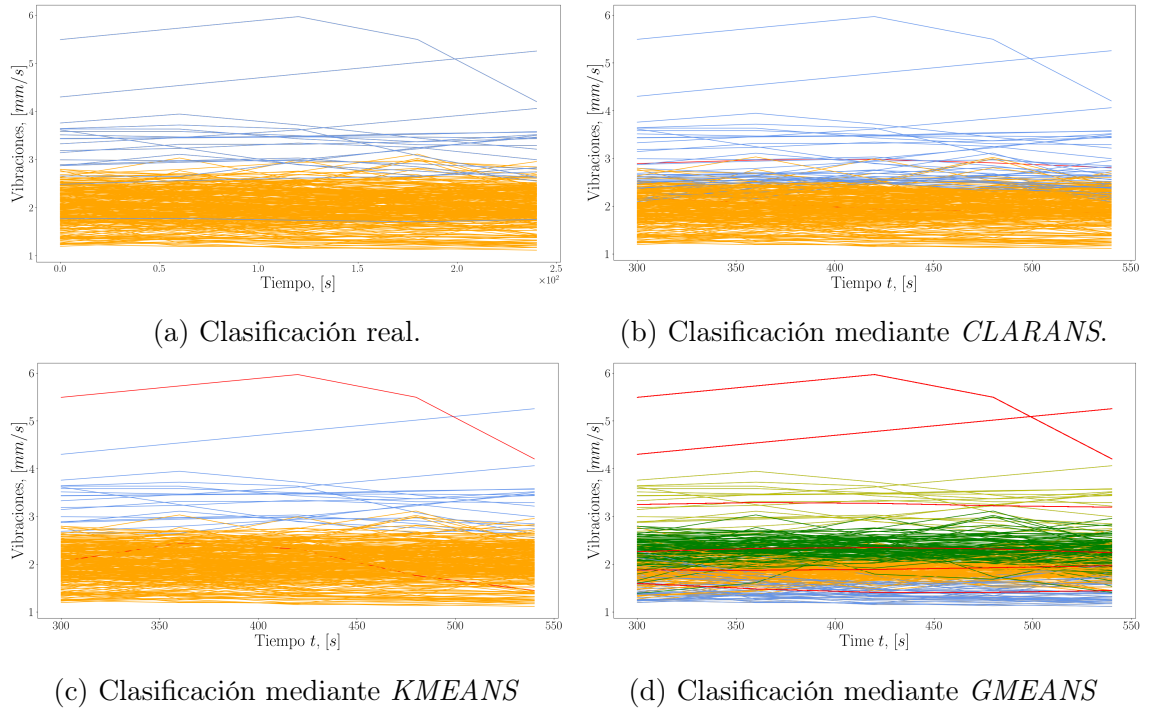


Figura 5.7: Clasificación de las curvas de velocidad introduciendo un factor de error de 1.5 sobre las primeras 20 curvas.

En las Figuras 5.6 - 5.8 se puede observar que para un factor de error igual a 2 (exceptuando con *GMEANS*, que clasifica las curvas en demasiados *clusters*), la clasificación es perfecta mientras que para factores de 1.5 y 0.5 se comienzan a producir fallos.

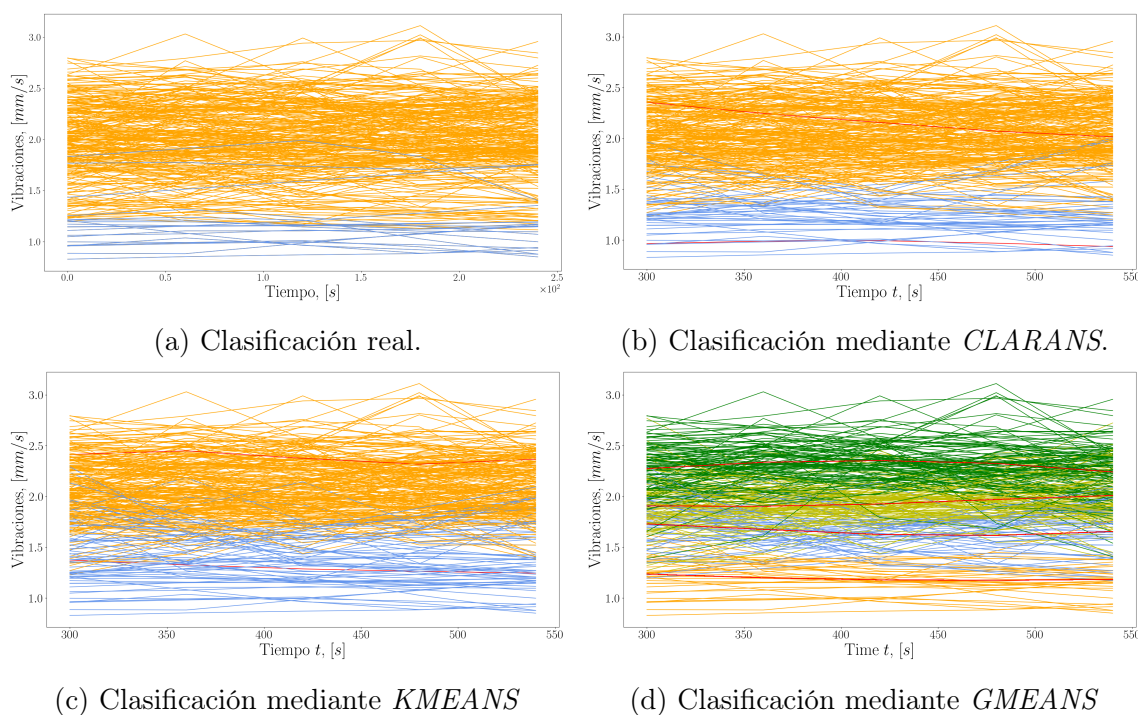


Figura 5.8: Clasificación de las curvas de velocidad introduciendo un factor de error de 0.5 sobre las primeras 20 curvas.

5.1. Clasificación de nuevos datos experimentales

Con el objetivo de analizar si el compresor está funcionando correctamente a partir de una base de datos previamente clasificada, bastará introducir las nuevas curvas y comprobar en qué grupo las clasifica el algoritmo sin tener que reclasificar la base de datos completa. Si consideramos como tipo 1 aquellas curvas a las que no hemos introducido error y tipo 2 a aquellas que hemos multiplicado por un factor, entonces si las nuevas curvas son clasificadas como tipo 1 podemos garantizar que el funcionamiento es correcto, mientras que serán de tipo 2 en el caso de un funcionamiento defectuoso. Bastaría establecer a partir de qué factor un fallo se considera aceptable.

En [14], se caracterizan las condiciones en las que puede trabajar una máquina industrial:

- A: buenas condiciones.
- B: condiciones aceptables para funcionar de forma permanente.
- C: condiciones aceptables a corto plazo hasta reparar el problema.
- D: condiciones inaceptables para el funcionamiento.

Por su parte, las máquinas se clasifican en 6 grupos principales para su análisis vibracional según la norma ISO 2372 y 2374:

- Grupo 1: Máquinas industriales pequeñas (motores y generadores) con alimentación de hasta 15kW.
- Grupo 2: Máquinas de tamaño medio (motores eléctricos) con alimentación entre 15 y 75 kW.
- Grupo 3: Aparatos grandes con partes rotativas sobre bases sólidas (turbocompresores).
- Grupo 4: Aparatos grandes con partes rotativas y conexiones flexibles, donde la frecuencia de las partes rotativas es mayor que la propia frecuencia del sistema.
- Grupo 5: Aparatos grandes con únicamente partes rotativas flexibles donde la frecuencia de las partes rotativas es mayor que la propia frecuencia del sistema.
- Grupo 6: Máquinas flexibles con fuerza de interferencia de desequilibrio y máquinas con acoplamientos libres (centrifugadoras, molinos, clasificadores, etc.)

En la Figura 5.1 se muestra una tabla mediante la cual se puede clasificar el estado de la máquina según la velocidad efectiva de sus vibraciones.

Considerando que el aparato se puede clasificar dentro del grupo dos y dado que la velocidad media de las vibraciones captadas mediante el vibrómetro situado en la parte delantera del compresor es de 2mm/s el funcionamiento del compresor es aceptable. Por tanto, para comenzar a apreciar un comportamiento de tipo C, en el que el aparato puede funcionar a corto plazo pero debe ser revisado, estas vibraciones medias deben ser superiores a 2.8.

Para imitar una situación similar en esta ocasión, se multiplicaran las primeras 150 curvas por 2, obteniendo una velocidad de vibración media

Intensidad de vibración		Grupo					
Puntuacion	Velocidad efectiva (mm/s)	1	2	3	4	5	6
0.71	0.45 a 0.71						
1.12	0.71 a 1.12					A	
1.8	1.12 a 1.8						
2.8	1.8 a 2.8						
4.5	2.8 a 4.5				B		
7.1	4.5 a 7.1						
11.2	7.1 a 11.2			C			
18	11.2 a 18						
28	18 a 28						
45	28 a 45		D				
71	45 a 71						

Tabla 5.1: Clasificación del estado de la maquinaria industrial según la velocidad efectiva de sus vibraciones. (Imagen de [14])

de 3mm/s . Como se puede apreciar en la Figura 5.9, la clasificación es en general correcta (exceptuando *GMEANS* que, de nuevo, emplea demasiados *clusters*).

Finalmente, es de interés comprobar si el algoritmo es capaz de clasificar nuevas curvas de vibraciones a partir de la clasificación que acaba de realizar. Para ello emplearemos el método *KMEDOIS*, ya que nos permite predecir a qué *cluster* pertenece una nueva curva dada. Con esta finalidad se han generado cuatro curvas con un comportamiento similar a las de la base de datos considerada. Dos de ellas se generan de manera que se asocian a un buen funcionamiento y las otras dos con un mal funcionamiento.

Se consideran dos grupos, el *cluster* 1 (en naranja), el que agrupa las vibraciones que el aparato presenta cuando tiene un comportamiento normal, y el *cluster* 2 (en azul), en el que se encuentran las curvas relacionadas con un comportamiento defectuoso. Si se introducen nuevos datos de las vibraciones captadas en un cierto periodo de tiempo y éstos se clasifican en el *cluster* 2 se puede concluir que el aparato precisa una revisión.

En la Figura 5.10 se pueden ver en magenta las curvas que han sido clasificadas en el *cluster* 1 y en verde las que lo han sido en el *cluster* 2. Como se puede observar, el algoritmo puede predecir correctamente a qué *cluster* pertenece cada curva. Todas las nuevas curvas se han clasificado correctamente.

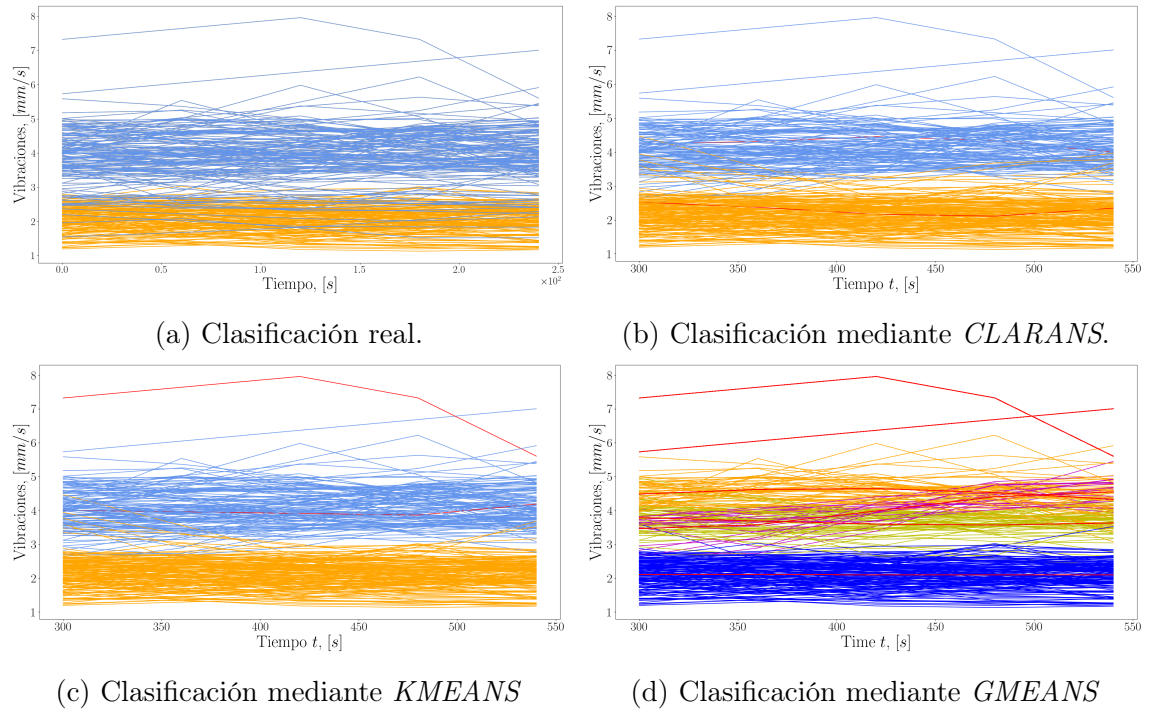


Figura 5.9: Clasificación de las curvas de vibraciones introduciendo un factor de error de 2 en las 150 primeras curvas.

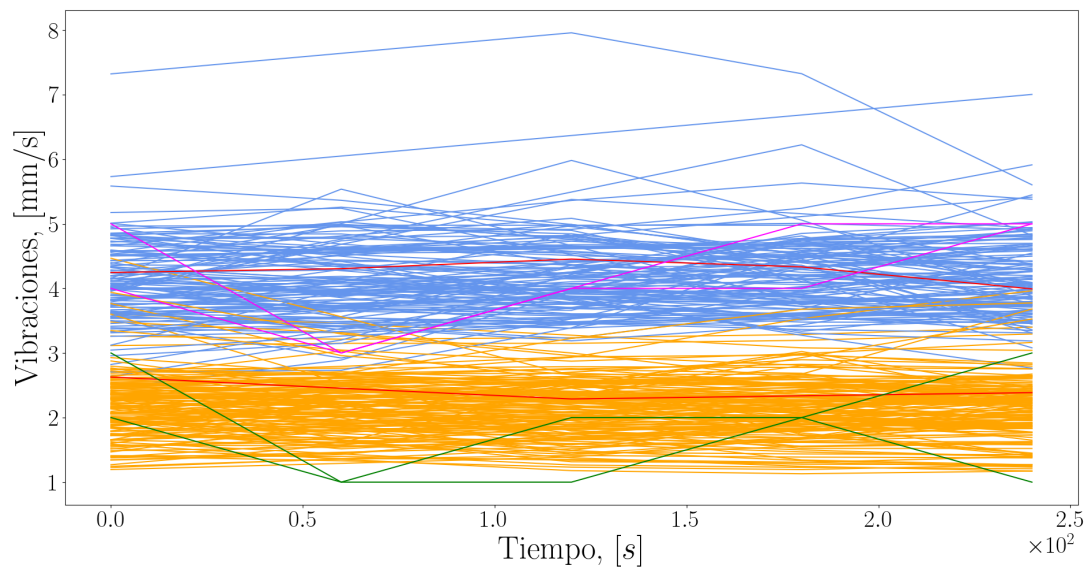


Figura 5.10: Predicción realizada por KMEDOIDS. En verde se representan las curvas que *KMEDOIS* clasifica en el *cluster* 1 (mal funcionamiento) y en magenta las que clasifica en el *cluster* 2 (buen funcionamiento).

Capítulo 6

Conclusiones y futuro trabajo

A pesar de las limitaciones temporales y la falta de datos de mal funcionamiento del compresor el presente trabajo ha puesto en manifiesto que es posible determinar el buen o mal funcionamiento del aparato usando sus vibraciones. Con todo, los resultados obtenidos están lejos de poder ser utilizados directamente por la empresa y requieren de más desarrollo.

En primer lugar, sería necesario verificar que, en efecto, el compresor pertenece al grupo 2 según la norma ISO 2372 y 2374 como hemos supuesto en el Capítulo 5. Por otra parte debería verificarse que tanto esta norma ISO 2372 y 2374 como los datos de la Figura 5.1 están actualizados o, por el contrario, es necesario realizar modificaciones. Una vez comprobado resta constatar que una situación de mal funcionamiento del compresor se corresponde con los resultados esperados. Cabe señalar que para que esto suceda sería necesario realizar un seguimiento a largo plazo del compresor, al menos hasta que el aparato deje de funcionar por una avería, llegados a este punto se dispondrán de medidas de velocidades de vibración en todas las etapas del fallo.

Conseguir identificar una posible avería antes de que esta suceda mediante las vibraciones del compresor ya es un gran logro, sin embargo esto puede abrirnos las puertas a nuevas preguntas: ¿Se puede detectar en qué componente del compresor se va a producir la avería? ¿Sucede de manera análoga en todos los compresores? ¿Se podría estimar en cuánto tiempo se producirá el fallo en un compresor? La respuesta a todas estas preguntas pueden ser de gran interés para la empresa, pues le permitiría ofrecer un mejor servicio y reducir los costes de reparación. Está claro que para poder responderlas es necesario profundizar en el funcionamiento de los distintos compresores y disponer de una enorme cantidad de datos de vibraciones.

Por otra parte, en el Capítulo 3 se han descrito distintos métodos de clasificación (estos son *CLARANS*, *KMEANS* y *GMEANS*) pero no se han detallado las ventajas y desventajas de cada uno de ellos. Sería oportuno determinar qué método es más eficiente para satisfacer los objetivos de la empresa. Para ello se debería determinar qué método optimiza simultáneamente la precisión y el tiempo de cálculo, lo que conllevaría un estudio en profundidad de los algoritmos.

Además, en el Capítulo 5 sólo se han realizado pruebas multiplicando por un mismo factor de error curvas consecutivas. Podría ser de interés realizar las pruebas multiplicando por distintos factores las curvas, en particular, haciendo que este factor aumente progresivamente con el tiempo. Esto podría aproximarse más a lo que sucede realmente con los compresores, de manera similar al planteamiento en la Sección 4.4 podría argumentarse que el aumento del desgaste de las piezas con el tiempo, (véase la Figura 4.18a), está asociado a mayores velocidades de vibración.

Bibliografía

- [1] Asurnipal. *Own work*, CC BY-SA 4.0. URL: <https://commons.wikimedia.org/w/index.php?curid=78991914>.
- [2] *Condition Monitoring and Prognostics Using Vibration Signals*. URL: <https://es.mathworks.com/help/predmaint/ug/condition-monitoring-and-prognostics-using-vibration-signals.html>.
- [3] Imad Dabbura. *K-means Clustering - Algorithm, Applications, Evaluation Methods, and Drawbacks*. 2018. URL: <https://imaddabbura.github.io/post/kmeans-clustering/>.
- [4] Yash Dagli. *Partitional Clustering using CLARANS Method with Python Example*. 2019. URL: <https://medium.com/analytics-vidhya/partitional-clustering-using-clarans-method-with-python-example-545dd84e58b4>.
- [5] Hamerly, Greg and Elkan y Charles. “Learning the K in K -Means”. En: *Advances in Neural Information Processing Systems* 17 (2004).
- [6] Leila Bakhtiaryfard; Shi-Xian Chen; Yu-Ren Wu; Sheng-Hung Hsieh; Yu-Hua Huang. *Vibration Analysis of Twin-Screw Compressors Under Partial Load Design: A Case Study*. Inf. téc. 2541. Purdue University, 2018. URL: <https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=3540&context=icec>.
- [7] Peter A. W. Lewis. “Distribution of the Anderson-Darling Statistic”. En: *The Annals of Mathematical Statistics* 32.4 (1961), págs. 1118-1124.
- [8] Raymond Ng y Jiawei Han. “CLARANS: A method for clustering objects for spatial data mining”. En: *Knowledge and Data Engineering, IEEE Transactions on* 14 (2002), págs. 1003-1016.
- [9] Andrei Novikov. “PyClustering: Data Mining Library”. En: *Journal of Open Source Software* 4.36 (2019), pág. 1230.

- [10] Pelleg Dan y Moore Andrew. *X-means: Extending K-means with Efficient Estimation of the Number of Clusters*. Inf. téc. Carnegie Mellon University, 2002. URL: <https://www.cs.cmu.edu/~dpelleg/download/xmeans.pdf>.
- [11] Peter Lancaster y Kestutis Salkauskas. *Curve and Surface Fitting. An Introduction*. University of Calgary: Academic Press, 1986.
- [12] *Sitio web de COFRICO*. URL: <https://www.cofrico.com/servicios/mantenimiento-predictivo-eficiencia-energetica/>.
- [13] Cheng Xiang You. *A theoretical study of rotor forces and torques in helical twin screw compressors*. Inf. téc. Division of Dynamics y Control Department of Mechanical Engineering. University of Strathclyde. Glasgow. UK, 1994. URL: <https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.249828>.
- [14] D. Hübbel; P. Zitek. “Screw compressor analysis from a vibration point-of-view”. En: *AIP Conference Proceedings* 1889 (2017).